CHAPTER

13

Adverse Selection, Signaling, and Screening

13.A Introduction

One of the implicit assumptions of the fundamental welfare theorems is that the characteristics of all commodities are observable to all market participants. Without this condition, distinct markets cannot exist for goods having differing characteristics, and so the complete markets assumption cannot hold. In reality, however, this kind of information is often asymmetrically held by market participants. Consider the following three examples:

(i) When a firm hires a worker, the firm may know less than the worker does about the worker's innate ability.
(ii) When an automobile insurance company insures an individual, the individual may know more than the company about her inherent driving skill and hence about her probability of having an accident.
(iii) In the used-car market, the seller of a car may have much better information about her car's quality than a prospective buyer does.

A number of questions immediately arise about these settings of *asymmetric information*: How do we characterize market equilibria in the presence of asymmetric information? What are the properties of these equilibria? Are there possibilities for welfare-improving market intervention? In this chapter, we study these questions, which have been among the most active areas of research in microeconomic theory during the last twenty years.

We begin, in Section 13.B, by introducing asymmetric information into a simple competitive market model. We see that in the presence of asymmetric information, market equilibria often fail to be Pareto optimal. The tendency for inefficiency in these settings can be strikingly exacerbated by the phenomenon known as *adverse selection*. Adverse selection arises when an informed individual's trading decisions depend on her privately held information in a manner that adversely affects uninformed market participants. In the used-car market, for example, an individual is more likely to decide to sell her car when she knows that it is not very good. When adverse selection is present, uninformed traders will be wary of any informed trader who wishes to trade with them, and their willingness to pay for the product offered will be low. Moreover, this fact may even further exacerbate the adverse selection problem: If the price that can be received by selling a used car is very low, only sellers with *really* bad cars will offer them for sale. As a result, we may see little trade in markets in which adverse selection is present, even if a great deal of trade would occur were information symmetrically held by all market participants.

We also introduce and study in Section 13.B an important concept for the analysis of market intervention in settings of asymmetric information: the notion of a *constrained Pareto optimal allocation.* These are allocations that cannot be Pareto improved upon by a central authority who, like market participants, cannot observe individuals' privately held information. A Pareto-improving market intervention can be achieved by such an authority only when the equilibrium allocation fails to be a constrained Pareto optimum. In general, the central authority's inability to observe individuals' privately held information leads to a more stringent test for Pareto-improving market intervention.

In Sections 13.C and 13.D, we study how market behavior may adapt in response to these informational asymmetries. In Section 13.C, we consider the possibility that informed individuals may find ways to *signal* information about their unobservable knowledge through observable actions. For example, a seller of a used car could offer to allow a prospective buyer to take the car to a mechanic. Because sellers who have good cars are more likely to be willing to take such an action, this offer can serve as a signal of quality. In Section 13.D, we consider the possibility that uninformed parties may develop mechanisms to distinguish, or *screen*, informed individuals who have differing information. For example, an insurance company may offer two policies: one with no deductible at a high premium and another with a significant deductible at a much lower premium. Potential insureds then *self-select*, with high-ability drivers choosing the policy with a deductible and low-ability drivers choosing the no-deductible policy. In both sections, we consider the welfare characteristics of the resulting market equilibria and the potential for Pareto-improving market intervention.

For expositional purposes, we present all the analysis that follows in terms of the labor market example (i). We should nevertheless emphasize the wide range of settings and fields within economics in which these issues arise. Some of these examples are developed in the exercises at the end of the chapter.

13.B Informational Asymmetries and Adverse Selection

Consider the following simple labor market model adapted from Akerlof's (1970) pioneering work:[1] there are many identical potential firms that can hire workers. Each produces the same output using an identical constant returns to scale technology in which labor is the only input. The firms are risk neutral, seek to maximize their expected profits, and act as price takers. For simplicity, we take the price of the firms' output to equal 1 (in units of a numeraire good).

Workers differ in the number of units of output they produce if hired by a firm,

1. Akerlof (1970) used the example of a used-car market in which only the seller of a used car knows if the car is a "lemon." For this reason, this type of model is sometimes referred to as a *lemons* model.

which we denote by θ.[2] We let $[\underline{\theta}, \bar{\theta}] \subset \mathbb{R}$ denote the set of possible worker productivity levels, where $0 \leq \underline{\theta} < \bar{\theta} < \infty$. The proportion of workers with productivity of θ or less is given by the distribution function $F(\theta)$, and we assume that $F(\cdot)$ is nondegenerate, so that there are at least two types of workers. The total number (or, more precisely, measure) of workers is N.

Workers seek to maximize the amount that they earn from their labor (in units of the numeraire good). A worker can choose to work either at a firm or at home, and we suppose that a worker of type θ can earn $r(\theta)$ on her own through home production. Thus, $r(\theta)$ is the opportunity cost to a worker of type θ of accepting employment; she will accept employment at a firm if and only if she receives a wage of at least $r(\theta)$ (for convenience, we assume that she accepts if she is indifferent).[3]

As a point of comparison, consider first the competitive equilibrium arising in this model when workers' productivity levels are *publicly observable*. Because the labor of each different type of worker is a distinct good, there is a distinct equilibrium wage $w^*(\theta)$ for each type θ. Given the competitive, constant returns nature of the firms, in a competitive equilibrium we have $w^*(\theta) = \theta$ for all θ (recall that the price of their output is 1), and the set of workers accepting employment in a firm is $\{\theta : r(\theta) \leq \theta\}$.[4]

As would be expected from the first fundamental welfare theorem, this competitive outcome is Pareto optimal. To verify this, recall that any Pareto optimal allocation of labor must maximize aggregate surplus (see Section 10.E). Letting $I(\theta)$ be a binary variable that equals 1 if a worker of type θ works for a firm and 0 otherwise, the sum of the aggregate surplus in these labor markets is equal to

$$\int_{\underline{\theta}}^{\bar{\theta}} N[I(\theta)\theta + (1 - I(\theta))r(\theta)]\, dF(\theta). \tag{13.B.1}$$

(This is simply the total revenue generated by the workers' labor.)[5] Aggregate surplus is therefore maximized by setting $I(\theta) = 1$ for those θ with $r(\theta) \leq \theta$ and $I(\theta) = 0$ otherwise (we again resolve indifference in favor of working at a firm). Put simply,

2. A worker's productivity could be random without requiring any change in the analysis that follows; in this case, θ is her *expected* (in a statistical sense) level of productivity.

3. An equivalent model arises from instead specifying $r(\theta)$ as the disutility of labor. In this alternative model, a worker of type θ has quasilinear preferences of the form $u(m, I) = m - r(\theta)I$, where m is the worker's consumption of the numeraire good and $I \in \{0, 1\}$ is a binary variable with $I = 1$ if the worker works and $I = 0$ if not. With these preferences, a worker again accepts employment if and only if she receives a wage of at least $r(\theta)$, and the rest of our analysis remains unaltered.

4. More precisely, there are also competitive equilibria in which $w^*(\theta) = \theta$ for all types of workers who are employed in the equilibrium [those with $r(\theta) \leq \theta$] and $w^*(\theta) \geq \theta$ for those types who are not [those with $r(\theta) > \theta$]. However, for the sake of expositional simplicity, when discussing competitive equilibria that involve no trade in this section we shall restrict attention to equilibrium wages that are equal to workers' (expected) productivity.

5. In Section 10.E, the aggregate surplus from an allocation in a product market (where firms produce output) was written as consumers' direct benefits from consumption of the good less firms' total costs of production. Here, in a labor market setting, a firm's "cost" of employing a worker is the positive revenue it earns, and a worker receives a direct utility (exclusive of any wage payments) of 0 if she works for a firm and $r(\theta)$ if she does not. Hence, aggregate surplus in these markets is equal to firms' total revenues, $\int NI(\theta)\theta\, dF(\theta)$, plus consumers' total revenue from home production, $\int N(1 - I(\theta))r(\theta)\, dF(\theta)$.

since a type θ worker produces at least as much at a firm as at home if and only if $r(\theta) \leq \theta$, in any Pareto optimal allocation the set of workers who are employed by the firms must be $\{\theta: r(\theta) \leq \theta\}$.

We now investigate the nature of competitive equilibrium when workers' productivity levels are *unobservable* by the firms. We begin by developing a notion of competitive equilibrium for this environment with asymmetric information.

To do so, note first that when workers' types are not observable, the wage rate must be independent of a worker's type, and so we will have a single wage rate w for all workers. Consider, then, the supply of labor as a function of the wage rate w. A worker of type θ is willing to work for a firm if and only if $r(\theta) \leq w$. Hence, the set of worker types who are willing to accept employment at wage rate w is

$$\Theta(w) = \{\theta: r(\theta) \leq w\}. \tag{13.B.2}$$

Consider, next, the demand for labor as a function of w. If a firm believes that the average productivity of workers who accept employment is μ, its demand for labor is given by

$$z(w) = \begin{cases} 0 & \text{if } \mu < w \\ [0, \infty] & \text{if } \mu = w \\ \infty & \text{if } \mu > w. \end{cases} \tag{13.B.3}$$

Now, if worker types in set Θ^* are accepting employment offers in a competitive equilibrium, and if firms' beliefs about the productivity of potential employees correctly reflect the actual average productivity of the workers hired in this equilibrium, then we must have $\mu = E[\theta \mid \theta \in \Theta^*]$. Hence, (13.B.3) implies that the demand for labor can equal its supply in an equilibrium with a positive level of employment if and only if $w = E[\theta \mid \theta \in \Theta^*]$. This leads to the notion of a competitive equilibrium presented in Definition 13.B.1.

Definition 13.B.1: In the competitive labor market model with unobservable worker productivity levels, a *competitive equilibrium* is a wage rate w^* and a set Θ^* of worker types who accept employment such that

$$\Theta^* = \{\theta: r(\theta) \leq w^*\} \tag{13.B.4}$$

and

$$w^* = E[\theta \mid \theta \in \Theta^*]. \tag{13.B.5}$$

Condition (13.B.5) involves *rational expectations* on the part of the firms. That is, firms correctly anticipate the average productivity of those workers who accept employment in the equilibrium.

Note, however, that the expectation in (13.B.5) is not well defined when *no* workers are accepting employment in an equilibrium (i.e., when $\Theta^* = \varnothing$). In the discussion that follows, we assume for simplicity that in this circumstance each firm's expectation of potential employees' average productivity is simply the unconditional expectation $E[\theta]$, and we take $w^* = E[\theta]$ in any such equilibrium. (As discussed in footnote 4, we restrict attention to wages that equal workers' expected productivity in any no-trade equilibrium. See Exercise 13.B.5 for the consequences of altering the assumption that expected productivity is $E[\theta]$ when $\Theta^* = \varnothing$.)

Asymmetric Information and Pareto Inefficiency

Typically, a competitive equilibrium as defined in Definition 13.B.1 will fail to be Pareto optimal. To see this point in the simplest-possible setting, consider the case where $r(\theta) = r$ for all θ (every worker is equally productive at home) and suppose that $F(r) \in (0, 1)$, so that there are some workers with $\theta > r$ and some with $\theta < r$. In this setting, the Pareto optimal allocation of labor has workers with $\theta \geq r$ accepting employment at a firm and those with $\theta < r$ not doing so.

Now consider the competitive equilibrium. When $r(\theta) = r$ for all θ, the set of workers who are willing to accept employment at a given wage, $\Theta(w)$, is either $[\underline{\theta}, \bar{\theta}]$ (if $w \geq r$) or $\varnothing$ (if $w < r$). Thus, $E[\theta \mid \theta \in \Theta(w)] = E[\theta]$ for all w and so by (13.B.5) the equilibrium wage rate must be $w^* = E[\theta]$. If $E[\theta] \geq r$, then *all* workers accept employment at a firm; if $E[\theta] < r$, then none do. Which type of equilibrium arises depends on the relative fractions of good and bad workers. For example, if there is a high fraction of low-productivity workers then, because firms cannot distinguish good workers from bad, they will be unwilling to hire any workers at a wage rate that is sufficient to have them accept employment (i.e., a wage of at least r). On the other hand, if there are very few low-productivity workers, then the average productivity of the workforce will be above r, and so the firms will be willing to hire workers at a wage that they are willing to accept. In one case, too many workers are employed relative to the Pareto optimal allocation, and in the other too few.

The cause of this failure of the competitive allocation to be Pareto optimal is simple to see: because firms are unable to distinguish among workers of differing productivities, the market is unable to allocate workers efficiently between firms and home production.[6]

Adverse Selection and Market Unraveling

A particularly striking breakdown in efficiency can arise when $r(\theta)$ varies with θ. In this case, the average productivity of those workers who are willing to accept employment in a firm depends on the wage, and a phenomenon known as *adverse selection* may arise. Adverse selection is said to occur when an informed individual's trading decision depends on her unobservable characteristics in a manner that adversely affects the uninformed agents in the market. In the present context, adverse selection arises when only relatively less capable workers are willing to accept a firm's employment offer at any given wage.

Adverse selection can have a striking effect on market equilibrium. For example, it may seem from our discussion of the case in which $r(\theta) = r$ for all θ that problems arise for the Pareto optimality of competitive equilibrium in the presence of asymmetric information only if there are some workers who should work for a firm and some who should not (since when either $\bar{\theta} < r$ or $\underline{\theta} > r$ the competitive equilibrium outcome is Pareto optimal). In fact, because of adverse selection, this is

6. Another way to understand the difficulty here is that asymmetric information leads to a situation with missing markets and thereby creates externalities (recall Chapter 11). When a worker of type $\theta > E[\theta] = w$ marginally reduces her supply of labor to a firm here, the firm is made worse off, in contrast with the situation in a competitive market with perfect information, where the wage exactly equals a worker's marginal productivity.

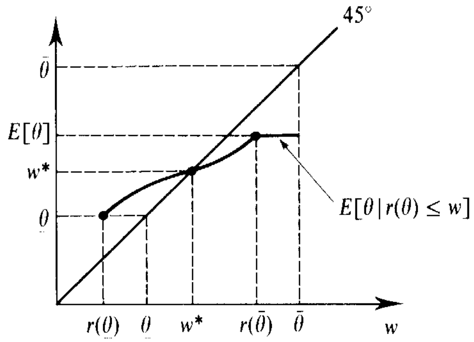


Figure 13.B.1
A competitive equilibrium with adverse selection.

not so; indeed, the market may fail completely despite the fact that *every* worker type should work at a firm.

To see the power of adverse selection, suppose that $r(\theta) \leq \theta$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and that $r(\cdot)$ is a strictly increasing function. The first of these assumptions implies that the Pareto optimal labor allocation has every worker type employed by a firm. The second assumption says that workers who are more productive at a firm are also more productive at home. It is this assumption that generates adverse selection: Because the payoff of home production is greater for more capable workers, only less capable workers accept employment at any given wage w [i.e., those with $r(\theta) \leq w$].

The expected value of worker productivity in condition (13.B.5) now depends on the wage rate. As the wage rate increases, more productive workers become willing to accept employment at a firm, and the average productivity of those workers accepting employment rises. For simplicity, from this point on, we assume that $F(\cdot)$ has an associated density function $f(\cdot)$, with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. This insures that the average productivity of those workers willing to accept employment, $E[\theta \mid r(\theta) \leq w]$, varies continuously with the wage rate on the set $w \in [r(\underline{\theta}), \infty]$.

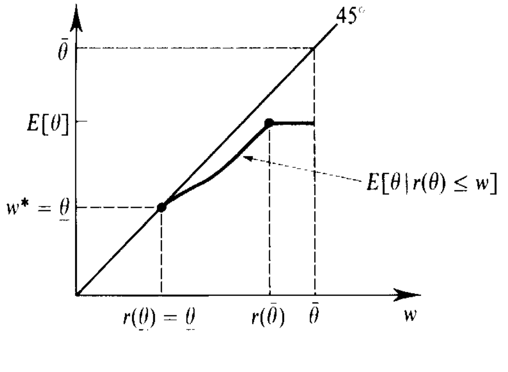
To determine the equilibrium wage, we use conditions (13.B.4) and (13.B.5). Together they imply that the competitive equilibrium wage w^* must satisfy

$$w^* = E[\theta \mid r(\theta) \leq w^*]. \tag{13.B.6}$$

We can use Figure 13.B.1 to study the determination of the equilibrium wage w^*. There we graph the values of $E[\theta \mid r(\theta) \leq w]$ as a function of w. This function gives the expected value of θ for workers who would choose to work for a firm when the prevailing wage is w. It is increasing in the level w for wages between $r(\underline{\theta})$ and $r(\bar{\theta})$, has a minimum value of $\underline{\theta}$ when $w = r(\underline{\theta})$, and attains a maximum value of $E[\theta]$ for $w \geq r(\bar{\theta})$.[7] The competitive equilibrium wage w^* is found by locating the wage rate at which this function crosses the 45-degree line; at this point, condition (13.B.6) is satisfied. The set of workers accepting employment at a firm is then $\Theta^* = \{\theta \colon r(\theta) \leq w^*\}$. Their average productivity is exactly w^*.[8]

7. The figure does not depict this function for wages below $r(\underline{\theta})$. Because $E[\theta] > r(\underline{\theta})$ in this model, no wage below $r(\underline{\theta})$ can be an equilibrium wage under our assumption that $E[\theta \mid \Theta(w) = \varnothing] = E[\theta]$.

8. For another diagrammatic determination of equilibrium, see Exercise 13.B.1.



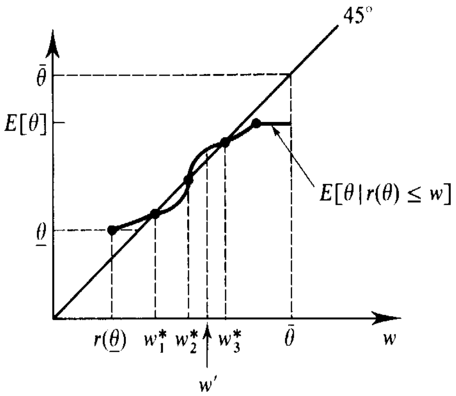


Figure 13.B.2 (left) Complete market failure.

Figure 13.B.3 (right) Multiple competitive equilibria.

We can see immediately from Figure 13.B.1 that the market equilibrium need not be efficient. The problem is that to get the best workers to accept employment at a firm, we need the wage to be at least $r(\bar{\theta})$. But in the case depicted, firms cannot break even at this wage because their inability to distinguish among different types of workers leaves them receiving only an expected output of $E[\theta] < r(\bar{\theta})$ from each worker that they hire. The presence of enough low-productivity workers therefore forces the wage down below $r(\bar{\theta})$, which in turn drives the best workers out of the market. But once the best workers are driven out of the market, the average productivity of the workforce falls, thereby further lowering the wage that firms are willing to pay. As a result, once the best workers are driven out of the market, the next-best may follow; the good may then be driven out by the mediocre.

How far can this process go? Potentially *very* far. To see this, consider the case depicted in Figure 13.B.2, where we have $r(\underline{\theta}) = \underline{\theta}$ and $r(\theta) < \theta$ for all other θ. There the equilibrium wage rate is $w^* = \underline{\theta}$, and only type $\underline{\theta}$ workers accept employment in the equilibrium. Because of adverse selection, essentially *no* workers are hired by firms (more precisely, a set of measure zero) even though the social optimum calls for *all* to be hired![9]

Example 13.B.1: To see an explicit example in which the market completely unravels let $r(\theta) = \alpha\theta$, where $\alpha < 1$, and let θ be distributed uniformly on [0, 2]. Thus, $r(\underline{\theta}) = \underline{\theta}$ (since $\underline{\theta} = 0$), and $r(\theta) < \theta$ for $\theta > 0$. In this case, $E[\theta \mid r(\theta) \leq w] = (w/2\alpha)$. For $\alpha > \frac{1}{2}$, $E[\theta \mid r(\theta) \leq 0] = 0$ and $E[\theta \mid r(\theta) \leq w] < w$ for all $w > 0$, as in Figure 13.B.2.[10]

The competitive equilibrium defined in Definition 13.B.1 need not be unique. Figure 13.B.3, for example, depicts a case in which there are three equilibria with strictly positive employment levels. Multiple competitive equilibria can arise because there is virtually no restriction on the slope of the function $E[\theta \mid r(\theta) \leq w]$. At any wage w, this slope depends on the density of workers who are just indifferent about accepting employment and so it can vary greatly if this density varies.

9. In this equilibrium, every agent receives the same payoff as if the market were abolished: every firm earns zero and a worker of type θ earns $r(\theta)$ for all θ (including $\theta = \underline{\theta}$).

10. This example is essentially the one developed in Akerlof (1970). His example corresponds to the case $\alpha = \frac{2}{3}$.

Note that the equilibria in Figure 13.B.3 can be *Pareto ranked*. Firms earn zero profits in any equilibrium, and workers are better off if the wage rate is higher (those workers who do not accept employment are indifferent; all other workers are strictly better off). Thus, the equilibrium with the highest wage Pareto dominates all the others. The low-wage, Pareto-dominated equilibria arise because of a *coordination failure*: the wage is too low because firms expect that the productivity of workers accepting employment is poor and, at the same time, only bad workers accept employment precisely because the wage is low.

A Game-Theoretic Approach

The notion of competitive equilibrium that we have employed above is that used by Akerlof (1970). We might ask whether these competitive equilibria can be viewed as the outcome of a richer model in which firms *could* change their offered wages but choose not to in equilibrium.

The situation depicted in Figure 13.B.3 might give you some concern in this regard. For example, consider the equilibrium with wage rate w_2^*. In this equilibrium, a firm that experimented with small changes in its wage offer would find that a small increase in its wage, say to the level $w' > w_2^*$ depicted in the figure, would raise its profits because it would then attract workers with an average productivity of $E[\theta \mid r(\theta) \leq w'] > w'$. Hence, it seems unlikely that a model in which firms could change their offered wages would ever lead to this equilibrium outcome. Similarly, at the equilibrium involving wage w_1^*, a firm that understood the structure of the market would realize that it could earn a strictly positive profit by raising its offered wage to w'.

To be more formal about this idea, consider the following game-theoretic model: The underlying structure of the market [e.g., the distribution of worker productivities $F(\cdot)$ and the reservation wage function $r(\cdot)$] is assumed to be common knowledge. Market behavior is captured in the following two-stage game: In stage 1, two firms simultaneously announce their wage offers (the restriction to two firms is without loss of generality). Then, in stage 2, workers decide whether to work for a firm and, if so, which one. (We suppose that if they are indifferent among some set of firms, then they randomize among them with equal probabilities.)[11]

Proposition 13.B.1 characterizes the subgame perfect Nash equilibria (SPNEs) of this game for the adverse selection model in which $r(\cdot)$ is strictly increasing with $r(\theta) \leq \theta$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and $F(\cdot)$ has an associated density $f(\cdot)$ with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$.

Proposition 13.B.1: Let W^* denote the set of competitive equilibrium wages for the adverse selection labor market model, and let $w^* = \text{Max}\,\{w\colon w \in W^*\}$.

(i) If $w^* > r(\underline{\theta})$ and there is an $\varepsilon > 0$ such that $E[\theta \mid r(\theta) \leq w'] > w'$ for all $w' \in (w^* - \varepsilon, w^*)$, then there is a unique pure strategy SPNE of the two-stage game-theoretic model. In this SPNE, employed workers receive

11. Note that if there is a single type of worker with productivity θ, this model is simply the labor market version of the Bertrand model of Section 12.C and has an equilibrium wage equal to θ, the competitive wage.

a wage of w^*, and workers with types in the set $\Theta(w^*) = \{\theta: r(\theta) \leq w^*\}$ accept employment in firms.

(ii) If $w^* = r(\underline{\theta})$, then there are multiple pure strategy SPNEs. However, in every pure strategy SPNE each agent's payoff exactly equals her payoff in the highest-wage competitive equilibrium.

Proof: To begin, note that in any SPNE a worker of type θ must follow the strategy of accepting employment only at one of the highest-wage firms, and of doing so if and only if its wage is at least $r(\theta)$.[12] Using this fact, we can determine the equilibrium behavior of the firms. We do so for each of the two cases in turn.

(i) $w^* > r(\underline{\theta})$: Note, first, that in any SPNE both firms must earn exactly zero. To see this, suppose that there is an SPNE in which a total of M workers are hired at a wage $\bar{w}$ and in which the aggregate profits of the two firms are

$$\Pi = M(E[\theta \mid r(\theta) \leq \bar{w}] - \bar{w}) > 0.$$

Note that $\Pi > 0$ implies that $M > 0$, which in turn implies that $\bar{w} \geq r(\underline{\theta})$. In this case, the (weakly) less-profitable firm, say firm j, must be earning no more than $\Pi/2$. But firm j can earn profits of at least $M(E[\theta \mid r(\theta) \leq \bar{w} + \alpha] - \bar{w} - \alpha)$ by instead offering wage $\bar{w} + \alpha$ for $\alpha > 0$. Since $E[\theta \mid r(\theta) \leq w]$ is continuous in w, these profits can be made arbitrarily close to Π by choosing α small enough. Thus, firm j would be better off deviating, which yields a contradiction: we must therefore have $\Pi \leq 0$. Because neither firm can have strictly negative profits in an SPNE (a firm can always offer a wage of zero), we conclude that both firms must be earning exactly zero in any SPNE.

From this fact, we know that if $\bar{w}$ is the highest wage rate offered by either of the two firms in an SPNE, then either $\bar{w} \in W^*$ (i.e., it must be a competitive equilibrium wage rate) or $\bar{w} < r(\underline{\theta})$ (it must be so low that no workers accept employment). But suppose that $\bar{w} < w^* = \text{Max}\,\{w: w \in W^*\}$. Then either firm can earn strictly positive expected profits by deviating and offering any wage rate $w' \in (w^* - \varepsilon, w^*)$. We conclude that the highest wage rate offered must equal w^* in any SPNE.

Finally, we argue that both firms naming w^* as their wage, plus the strategies for workers described above, constitute an SPNE. With these strategies, both firms earn zero. Neither firm can earn a positive profit by unilaterally lowering its wage because it gets no workers if it does so. To complete the argument, we show that $E[\theta \mid r(\theta) \leq w] < w$ at every $w > w^*$, so that no unilateral deviation to a higher wage can yield a firm positive profits either. By hypothesis, w^* is the highest competitive wage. Hence, there is no $w > w^*$ at which $E[\theta \mid r(\theta) \leq w] = w$. Therefore, because $E[\theta \mid r(\theta) \leq w]$ is continuous in w, $E[\theta \mid r(\theta) \leq w] - w$ must have the same sign for all $w > w^*$. But we cannot have $E[\theta \mid r(\theta) \leq w] > w$ for all $w > w^*$ because, as $w \to \infty$, $E[\theta \mid r(\theta) \leq w] \to E[\theta]$, which, under our assumptions, is finite. We must therefore have $E[\theta \mid r(\theta) \leq w] < w$ at all $w > w^*$. This completes the argument for case (i).

The assumption that there exists an $\varepsilon > 0$ such that $E[\theta \mid r(\theta) \leq w'] > w'$ for all $w' \in (w^* - \varepsilon, w^*)$ rules out pathological cases such as that depicted in Figure 13.B.4.

(ii) $w^* = r(\underline{\theta})$: In this case, $E[\theta \mid r(\theta) \leq w] < w$ for all $w > w^*$, so that any firm attracting workers at a wage in excess of w^* incurs losses. Moreover, a firm must

12. Recall that we assume that a worker accepts employment whenever she is indifferent.

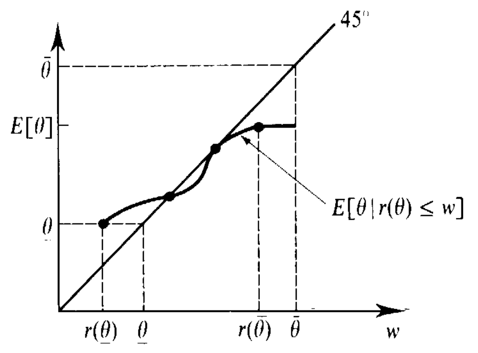


Figure 13.B.4
A pathological example.

earn exactly zero by announcing any $w \le w^*$. Hence, the set of wage offers (w_1, w_2) that can arise in an SPNE is $\{(w_1, w_2): w_j \le w^* \text{ for } j = 1, 2\}$. In every one of these SPNEs, all agents earn exactly what they earn at the competitive equilibrium involving wage rate w^*: both firms earn zero, and a worker of type θ earns $r(\theta)$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. ■

One difference between this game-theoretical model and the notion of competitive equilibrium specified in Definition 13.B.1 involves the level of firms' sophistication. In the competitive equilibria of Definition 13.B.1, firms can be fairly unsophisticated. They need know only the average productivity level of the workers who accept employment at the going equilibrium wage; they need not have any idea of the underlying market mechanism. In contrast, in the game-theoretic model, firms understand the entire structure of the market, including the full relationship that exists between the wage rate and the quality of employed workers. The game-theoretic model tells us that if sophisticated firms have the ability to make wage offers, then we break the coordination problem described above. If the wage is too low, some firm will find it in its interest to offer a higher wage and attract better workers; the highest-wage competitive outcome must then arise.[13]

Constrained Pareto Optima and Market Intervention

We have seen that the presence of asymmetric information often results in market equilibria that fail to be Pareto optimal. As a consequence, a central authority who knows all agents' private information (e.g., worker types in the models above), and can engage in lump-sum transfers among agents in the economy, can achieve a Pareto improvement over these outcomes.

In practice, however, a central authority may be no more able to observe agents' private information than are market participants. Without this information, the authority will face additional constraints in trying to achieve a Pareto improvement. For example, arranging lump-sum transfers among workers of different types will be impossible because the authority cannot observe workers' types directly. For Pareto-improving market intervention to be possible in this case, a more stringent test must therefore be passed. An allocation that cannot be Pareto improved by an

13. See Exercise 13.B.6, however, for an example of a model of adverse selection in which, for some parameter values, the highest-wage competitive equilibrium is *not* an SPNE of our game-theoretic model.

authority who is unable to observe agents' private information is known as a *constrained* (or *second-best*) *Pareto optimum.* Because it is more difficult to generate a Pareto improvement in the absence of an ability to observe agents' types, a constrained Pareto optimal allocation need not be (fully) Pareto optimal [however, a (full) Pareto optimum is necessarily a constrained Pareto optimum].

Here, as an example, we shall study whether Pareto-improving market intervention is possible in the context of our adverse selection model (where $r(\cdot)$ is strictly increasing with $r(\theta) \leq \theta$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and $F(\cdot)$ has an associated density $f(\cdot)$ with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$) when the central authority cannot observe worker types. That is, we study whether the competitive equilibria of this adverse selection model are constrained Pareto optima.

In general, the formal analysis of this problem uses tools that we develop in Section 14.C in our study of principal-agent models with hidden information (see, in particular, the discussion of monopolistic screening). As these techniques have yet to be introduced, we shall not analyze this problem fully here. (Once you have studied Section 14.C, however, refer back to the discussion in small type at the end of this section.) Nevertheless, we can convey much of the analysis here.

By way of motivation, note first that in examining whether a Pareto improvement relative to a market equilibrium is possible, we might as well simply think of intervention schemes in which the authority runs the firms herself and tries to achieve a Pareto improvement for the workers (the firms' owners will then earn exactly what they were earning in the equilibrium, namely zero profits). Second, because the authority cannot distinguish directly among different types of workers, any differences in lump-sum transfers to or from a worker can depend only on whether the worker is employed (the workers otherwise appear identical). Thus, intuitively, there should be no loss of generality in restricting attention to interventions in which the authority runs the firms herself, offers a wage of w_e to those accepting employment, an unemployment benefit of w_u to those who do not [these workers also receive $r(\theta)$], leaves the workers free to choose whether to accept employment in a firm, and balances her budget. (In the small-type discussion at the end of this section, we show formally that this is the case.)

Given this background, can the competitive equilibria of our adverse selection model be Pareto-improved upon in this way? Consider, first, dominated competitive equilibria, that is, competitive equilibria that are Pareto dominated by some other competitive equilibrium (e.g., the equilibrium with wage rate w_1^* shown in Figure 13.B.3). A central authority who is unable to observe worker types can always implement the best (highest-wage) competitive equilibrium outcome. She need only set $w_e = w^*$, the highest competitive equilibrium wage, and $w_u = 0$. All workers in set $\Theta(w^*)$ then accept employment in a firm and, since $w^* = E[\theta \mid r(\theta) \leq w^*]$, the authority exactly balances her budget.[14] Thus, the outcome in such an equilibrium is *not* a constrained Pareto optimum. In this case, the planner is essentially able to step in and solve the coordination failure that is keeping the market at the low-wage equilibrium.

14. An equivalent but less heavy-handed intervention would have the authority simply require any operating firm to pay a wage rate equal to w^*. Firms will be willing to remain operational because they break even at this wage rate, and a Pareto improvement results.

What about the highest-wage competitive equilibrium (i.e., the SPNE outcome in the game-theoretic model of Proposition 13.B.1)? As Proposition 13.B.2 shows, any such equilibrium *is* a constrained Pareto optimum in this model.

Proposition 13.B.2: In the adverse selection labor market model (where $r(\cdot)$ is strictly increasing with $r(\theta) \leq \theta$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and $F(\cdot)$ has an associated density $f(\cdot)$ with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$), the highest-wage competitive equilibrium is a constrained Pareto optimum.

Proof: If all workers are employed in the highest wage competitive equilibrium then the outcome is fully (and, hence, constrained) Pareto optimal. So suppose some are not employed. Note, first, that for any wage w_e and unemployment benefit w_u offered by the central authority the set of worker types accepting employment has the form $[\underline{\theta}, \hat{\theta}]$ for some $\hat{\theta}$ [it is $\{\theta: w_u + r(\theta) \leq w_e\}$]. Suppose, then, that the authority attempts to implement an outcome in which worker types $\theta \leq \hat{\theta}$ for $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$ accept employment. To do so, she must choose w_e and w_u so that

$$w_u + r(\hat{\theta}) = w_e. \tag{13.B.7}$$

In addition, to balance her budget, w_u and w_e must also satisfy[15]

$$w_e F(\hat{\theta}) + w_u(1 - F(\hat{\theta})) = \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta)\, d\theta. \tag{13.B.8}$$

Substituting into (13.B.7) from (13.B.8), we find that, given the choice of $\hat{\theta}$, the values of w_u and w_e must be

$$w_u(\hat{\theta}) = \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta)\, d\theta - r(\hat{\theta})F(\hat{\theta}) \tag{13.B.9}$$

and

$$w_e(\hat{\theta}) = \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta)\, d\theta + r(\hat{\theta})(1 - F(\hat{\theta})), \tag{13.B.10}$$

or, equivalently,

$$w_u(\hat{\theta}) = F(\hat{\theta})(E[\theta \mid \theta \leq \hat{\theta}] - r(\hat{\theta})) \tag{13.B.11}$$

$$w_e(\hat{\theta}) = F(\hat{\theta})(E[\theta \mid \theta \leq \hat{\theta}] - r(\hat{\theta})) + r(\hat{\theta}). \tag{13.B.12}$$

Now, let θ^* denote the highest worker type who accepts employment in the highest-wage competitive equilibrium. We know that $r(\theta^*) = E[\theta \mid \theta \leq \theta^*]$. Hence, from conditions (13.B.11) and (13.B.12), we see that $w_u(\theta^*) = 0$ and $w_e(\theta^*) = r(\theta^*)$. Thus, the outcome when the authority sets $\hat{\theta} = \theta^*$ is exactly the same as in the highest-wage competitive equilibrium.

We now examine whether a Pareto improvement can be achieved by setting $\hat{\theta} \neq \theta^*$. Note that for any $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$ with $\hat{\theta} \neq \theta^*$, type $\underline{\theta}$ workers are worse off than in the equilibrium if $w_e(\hat{\theta}) < r(\theta^*)$ [$r(\theta^*)$ is their wage in the equilibrium] and type $\bar{\theta}$ workers are worse off if $w_u(\hat{\theta}) < 0$.

Consider $\hat{\theta} < \theta^*$ first. Since $r(\theta^*) > r(\hat{\theta})$, condition (13.B.10) implies that

$$w_e(\hat{\theta}) \leq \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta)\, d\theta + r(\theta^*)(1 - F(\hat{\theta})),$$

15. The authority will never wish to run a budget surplus. If w_u and w_e lead to a budget surplus, then setting $\hat{w}_u = w_u + \varepsilon$ and $\hat{w}_e = w_e + \varepsilon$ for some $\varepsilon > 0$ is budget feasible and is Pareto superior. (Note that the set of workers accepting employment would be unchanged.)

and so

$$\begin{aligned} w_e(\hat{\theta}) - r(\theta^*) &\leq F(\hat{\theta})(E[\theta \mid \theta \leq \hat{\theta}] - r(\theta^*)) \\ &= F(\hat{\theta})(E[\theta \mid \theta \leq \hat{\theta}] - E[\theta \mid \theta \leq \theta^*]) \\ &< 0. \end{aligned}$$

Thus, type $\underline{\theta}$ workers must be made worse off by any such intervention.

Now consider $\hat{\theta} > \theta^*$. We know that $E[\theta \mid r(\theta) \leq w] < w$ for all $w > w^*$ (see the proof of Proposition 13.B.1). Thus, since $r(\theta^*) = w^*$ and $r(\cdot)$ is strictly increasing, we have $E[\theta \mid r(\theta) \leq r(\hat{\theta})] < r(\hat{\theta})$ for all $\hat{\theta} > \theta^*$. Moreover,

$$E[\theta \mid r(\theta) \leq r(\hat{\theta})] = E[\theta \mid \theta \leq \hat{\theta}],$$

and so $E[\theta \mid \theta \leq \hat{\theta}] - r(\hat{\theta}) < 0$ for all $\hat{\theta} > \theta^*$. But condition (13.B.11) then implies that $w_u(\hat{\theta}) < 0$ for all $\hat{\theta} > \theta^*$, and so type $\bar{\theta}$ workers are made worse off by any such intervention. ■

Hence, when a central authority cannot observe worker types, her options may be severely limited. Indeed, in the adverse selection model just considered, the authority is unable to create a Pareto improvement as long as the highest-wage competitive equilibrium (the SPNE outcome of the game-theoretic model of Proposition 13.B.1) is the market outcome.[16] More generally, whether Pareto-improving market intervention is possible in situations of asymmetric information depends on the specifics of the market under study (and as we have already seen, possibly on which equilibria result). Exercises 13.B.8 and 13.B.9 provide two examples of models in which the highest-wage competitive equilibrium may fail to be a constrained Pareto optimum.

Although it is impossible to Pareto improve a constrained Pareto optimal allocation, market intervention could still be justified in the pursuit of distributional aims. For example, if social welfare is given by the sum of weighted worker utilities

$$\int_{\underline{\theta}}^{\bar{\theta}} [I(\theta)\theta + (1 - I(\theta))r(\theta)]\lambda(\theta)\, dF(\theta), \tag{13.B.13}$$

where $\lambda(\theta) > 0$ for all θ, then social welfare may be increased even though some worker types end up worse off. In the applied literature, for example, it is common to see aggregate surplus used as the social welfare function, which is equivalent to the choice of $\lambda(\theta) = N$ for all θ.[17] When society has this social welfare function, social welfare can be raised relative to the competitive equilibrium in Figure 13.B.1 (which, by Proposition 13.B.2, is a constrained Pareto optimum) simply by mandating that all workers must work for a firm and that all firms must

16. Proposition 13.B.2 can also be readily generalized to allow $r(\theta) > \theta$ for some θ. (See Exercise 13.B.10.)

17. Note that when types cannot be observed, aggregate surplus is no longer a valid welfare measure for *any* social welfare function because, unlike the case of perfect information, lump-sum transfers across worker types are infeasible. (See Section 10.E for a discussion of the need for lump-sum transfers to justify aggregate surplus as a welfare measure for any social welfare function.)

pay workers a wage of $E(\theta)$. Although workers of type $\bar{\theta}$ are made worse off by this intervention, welfare as measured by aggregate surplus increases.[18]

An interesting interpretation of the choice of aggregate surplus as a social welfare function is in terms of an unborn worker's ex ante expected utility. In particular, imagine that each worker originally has a probability $f(\theta)$ of ending up a type θ worker. If this unborn worker is risk neutral, then her ex ante expected utility is exactly equal to expression (13.B.13) with $\lambda(\theta) = 1$ for all θ. Thus, maximization of aggregate surplus is equivalent to maximization of this unborn worker's expected utility. We might then say that an allocation is an *ex ante constrained Pareto optimum* in this model if, in the absence of an ability to observe worker types, it is impossible to devise a market intervention that raises aggregate surplus. We see, therefore, that whether an allocation is a constrained optimum (and, thus, whether a planned intervention leads to a Pareto improvement) can depend on the point at which the welfare evaluation is conducted (i.e., before the workers know their types, or after).[19]

Let us now use the techniques of Section 14.C to show formally that we can restrict attention in searching for a Pareto improvement to interventions of the type considered above. We shall look for a Pareto improvement for the workers keeping the profits of the firms' owners nonnegative. For notational simplicity, we shall treat the firms as a single aggregate firm.

By the revelation principle (see Section 14.C), we know that we can restrict attention to direct revelation mechanisms in which every worker type tells the truth. Here a direct revelation mechanism assigns, for each worker type $\theta \in [\underline{\theta}, \bar{\theta}]$, a payment from the authority to the worker of $w(\theta) \in \mathbb{R}$, a tax $t(\theta)$ paid by the firm to the authority, and an employment decision $I(\theta) \in \{0, 1\}$. The set of feasible mechanisms here are those that satisfy the *individual rationality constraint* for the firm,

$$\int_{\underline{\theta}}^{\bar{\theta}} [I(\theta)\theta - t(\theta)]\, dF(\theta) \geq 0, \tag{13.B.14}$$

the *budget balance condition* for the central authority,

$$\int_{\underline{\theta}}^{\bar{\theta}} [t(\theta) - w(\theta)]\, dF(\theta) \geq 0, \tag{13.B.15}$$

and the *truth-telling* (or *incentive compatibility*, or *self-selection*) *constraints* that say that for all θ and $\hat{\theta}$

$$w(\theta) + (1 - I(\theta))r(\theta) \geq w(\hat{\theta}) + (1 - I(\hat{\theta}))r(\theta). \tag{13.B.16}$$

Note, first, that mechanism $[w(\cdot), t(\cdot), I(\cdot)]$ is feasible only if $[w(\cdot), I(\cdot)]$ satisfies both condition (13.B.16) and

$$\int_{\underline{\theta}}^{\bar{\theta}} [I(\theta)\theta - w(\theta)]\, dF(\theta) \geq 0. \tag{13.B.17}$$

18. Moreover, because lump-sum transfers among different types of workers are not possible in the absence of an ability to observe worker types, the achievement of these distributional aims actually *requires* direct intervention in the labor market, in contrast with the case of perfect information.

19. Holmstrom and Myerson (1983) call this ex ante notion of constrained Pareto optimality *ex ante incentive efficiency*. Their terminology refers to the fact that we are taking an ex ante perspective in evaluating welfare (before the realization of worker types) and that a central authority who cannot observe worker types faces *incentive constraints* if she wants to induce workers to reveal their types. Holmstrom and Myerson call our previous notion of constrained Pareto efficiency *interim incentive efficiency* because the perspective used to assess Pareto optimality is that of workers who already know their types. See Section 23.F for a more general discussion of these concepts.

Moreover, if $[w(\cdot), I(\cdot)]$ satisfies (13.B.16) and (13.B.17), then there exists a $t(\cdot)$ such that $[w(\cdot), t(\cdot), I(\cdot)]$ satisfies (13.B.14)–(13.B.16). Condition (13.B.17), however, is exactly the budget constraint faced by a central authority who runs the firms herself. Hence, we can restrict attention to schemes in which the authority runs the firms herself and uses a direct revelation mechanism $[w(\cdot), I(\cdot)]$ satisfying (13.B.16) and (13.B.17).

Now consider any two types θ' and θ'' for which $I(\theta') = I(\theta'')$. Setting $\theta = \theta'$ and $\hat{\theta} = \theta''$ in condition (13.B.16), we see that we must have $w(\theta') \geq w(\theta'')$. Likewise, letting $\theta = \theta''$ and $\hat{\theta} = \theta'$, we must have $w(\theta'') \geq w(\theta')$. Together, this implies that $w(\theta') = w(\theta'')$. Since $I(\theta) \in \{0, 1\}$, we see that any feasible mechanism $[w(\cdot), I(\cdot)]$ can be viewed as a scheme that gives each worker a choice between two outcomes, $(w_e, I = 1)$ and $(w_u, I = 0)$ and satisfies the budget balance condition (13.B.17). This is exactly the class of mechanisms studied above.

13.C Signaling

Given the problems observed in Section 13.B, one might expect mechanisms to develop in the marketplace to help firms distinguish among workers. This seems plausible because both the firms and the high-ability workers have incentives to try to accomplish this objective. The mechanism that we examine in this section is that of *signaling*, which was first investigated by Spence (1973, 1974). The basic idea is that high-ability workers may have actions they can take to distinguish themselves from their low-ability counterparts.

The simplest example of such a signal occurs when workers can submit to some costless test that reliably reveals their type. It is relatively straightforward to show that in any subgame perfect Nash equilibrium all workers with ability greater than $\underline{\theta}$ will submit to the test and the market will achieve the full information outcome (see Exercise 13.C.1). Any worker who chooses not to take the test will be correctly treated as being no better than the worst type of worker.

However, in many instances, no procedure exists that directly reveals a worker's type. Nevertheless, as the analysis in this section reveals, the potential for signaling may still exist.

Consider the following adaptation of the model discussed in Section 13.B. For simplicity, we restrict attention to the case of two types of workers with productivities θ_H and θ_L, where $\theta_H > \theta_L > 0$ and $\lambda = \text{Prob}\,(\theta = \theta_H) \in (0, 1)$. The important extension of our previous model is that before entering the job market a worker can get some education, and the amount of education that a worker receives is observable. To make matters particularly stark, we assume that education does *nothing* for a worker's productivity (see Exercise 13.C.2 for the case of productive signaling). The cost of obtaining education level e for a type θ worker (the cost may be of either monetary or psychic origin) is given by the twice continuously differentiable function $c(e, \theta)$, with $c(0, \theta) = 0$, $c_e(e, \theta) > 0$, $c_{ee}(e, \theta) > 0$, $c_\theta(e, \theta) < 0$ for all $e > 0$, and $c_{e\theta}(e, \theta) < 0$ (subscripts denote partial derivatives). Thus, both the cost and the marginal cost of education are assumed to be lower for high-ability workers; for example, the work required to obtain a degree might be easier for a high-ability individual. Letting $u(w, e \mid \theta)$ denote the utility of a type θ worker who chooses education level e and receives wage w, we take $u(w, e \mid \theta)$ to equal her wage less any educational costs incurred: $u(w, e \mid \theta) = w - c(e, \theta)$. As in Section 13.B, a worker of type θ can earn $r(\theta)$ by working at home.

In the analysis that follows, we shall see that this otherwise useless education may serve as a signal of unobservable worker productivity. In particular, equilibria emerge in which high-productivity workers choose to get more education than low-productivity workers and firms correctly take differences in education levels as a signal of ability. The welfare effects of signaling activities are generally ambiguous. By revealing information about worker types, signaling can lead to a more efficient allocation of workers' labor, and in some instances to a Pareto improvement. At the same time, because signaling activity is costly, workers' welfare may be reduced if they are compelled to engage in a high level of signaling activity to distinguish themselves.

To keep things simple, throughout most of this section we concentrate on the special case in which $r(\theta_H) = r(\theta_L) = 0$. Note that under this assumption the unique equilibrium that arises in the absence of the ability to signal (analyzed in Section 13.B) has all workers employed by firms at a wage of $w^* = E[\theta]$ and is Pareto efficient. Hence, our study of this case emphasizes the potential inefficiencies created by signaling. After studying this case in detail, we briefly illustrate (in small type) how, with alternative assumptions about the function $r(\cdot)$, signaling may instead generate a Pareto improvement.

A portion of the game tree for this model is shown in Figure 13.C.1. Initially, a random move of nature determines whether a worker is of high or low ability. Then, conditional on her type, the worker chooses how much education to obtain. After obtaining her chosen education level, the worker enters the job market. Conditional on the observed education level of the worker, two firms simultaneously make wage offers to her. Finally, the worker decides whether to work for a firm and, if so, which one.

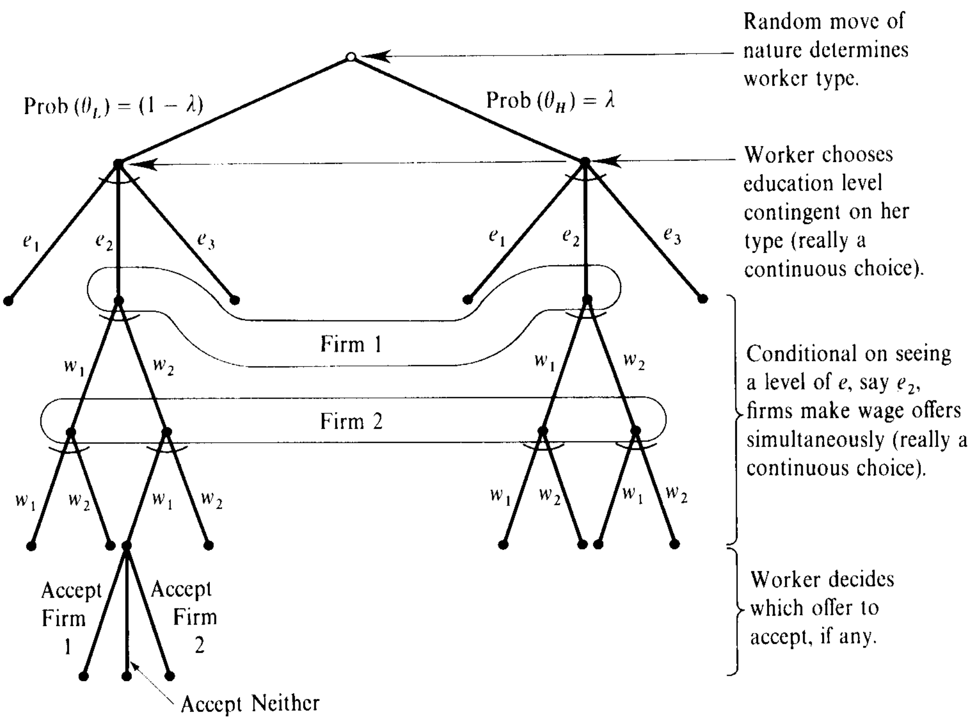


Figure 13.C.1
The extensive form of the education signaling game.

Note that, in contrast with the model of Section 13.B, here we explicitly model only a single worker of unknown type; the model with many workers can be thought of as simply having many of these single-worker games going on simultaneously, with the fraction of high-ability workers in the market being λ. In discussing the equilibria of this game, we often speak of the "high-ability workers" and "low-ability workers," having the many-workers case in mind.

The equilibrium concept we employ is that of a weak perfect Bayesian equilibrium (see Definition 9.C.3), but with an added condition. Put formally, we require that, in the game tree depicted in Figure 13.C.1, the firms' beliefs have the property that, for each possible choice of e, there exists a number $\mu(e) \in [0, 1]$ such that: (i) firm 1's belief that the worker is of type θ_H after seeing her choose e is $\mu(e)$ and (ii) after the worker has chosen e, firm 2's belief that the worker is of type θ_H and that firm 1 has chosen wage offer w is precisely $\mu(e)\sigma_1^*(w \mid e)$, where $\sigma_1^*(w \mid e)$ is firm 1's equilibrium probability of choosing wage offer w after observing education level e. This extra condition adds an element of commonality to the firms' beliefs about the type of worker who has chosen e, and requires that the firms' beliefs about each others' wage offers following e are consistent with the equilibrium strategies both on and off the equilibrium path.

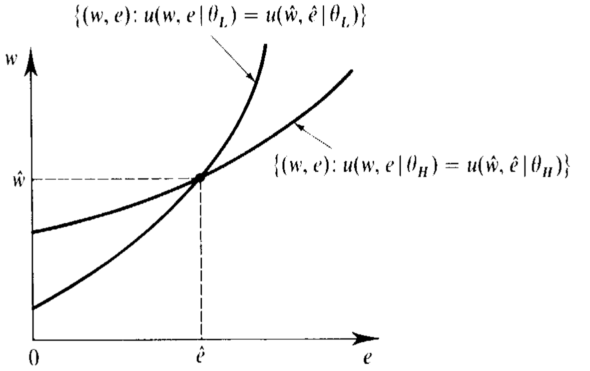
We refer to a weak perfect Bayesian equilibrium satisfying this extra condition on beliefs as a *perfect Bayesian equilibrium* (PBE). Fortunately, this PBE notion can more easily, and equivalently, be stated as follows: A set of strategies and a belief function $\mu(e) \in [0, 1]$ giving the firms' common probability assessment that the worker is of high ability after observing education level e is a PBE if

(i) The worker's strategy is optimal given the firm's strategies.
(ii) The belief function $\mu(e)$ is derived from the worker's strategy using Bayes' rule where possible.
(iii) The firms' wage offers following each choice e constitute a Nash equilibrium of the simultaneous-move wage offer game in which the probability that the worker is of high ability is $\mu(e)$.[20]

In the context of the model studied here, this notion of a PBE is equivalent to the sequential equilibrium concept discussed in Section 9.C. We also restrict our attention throughout to pure strategy equilibria.

We begin our analysis at the end of the game. Suppose that after seeing some education level e, the firms attach a probability of $\mu(e)$ that the worker is type θ_H. If so, the expected productivity of the worker is $\mu(e)\theta_H + (1 - \mu(e))\theta_L$. In a simultaneous-move wage offer game, the firms' (pure strategy) Nash equilibrium wage offers equal the worker's expected productivity (this game is very much like the Bertrand pricing game discussed in Section 12.C). Thus, in any (pure strategy) PBE, we must have both firms offering a wage exactly equal to the worker's expected productivity, $\mu(e)\theta_H + (1 - \mu(e))\theta_L$.

20. Thus, the extra condition we add imposes equilibrium-like play in parts of the tree off the equilibrium path. See Section 9.C for a discussion of the need to augment the weak perfect Bayesian equilibrium concept to achieve this end.



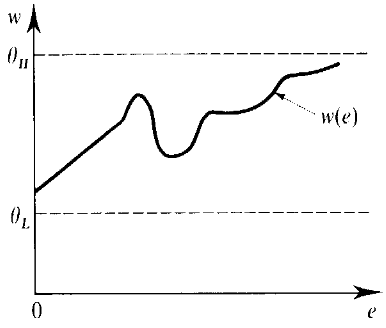


Figure 13.C.2 (left)
Indifference curves for high- and low-ability workers: the single-crossing property.

Figure 13.C.3 (right)
A wage schedule.

Knowing this fact, we turn to the issue of the worker's equilibrium strategy, her choice of an education level contingent on her type. As a first step in this analysis, it is useful to examine the worker's preferences over (wage rate, education level) pairs. Figure 13.C.2 depicts an indifference curve for each of the two types of workers (with wages measured on the vertical axis and education levels measured on the horizontal axis). Note that these indifference curves cross only once and that, where they do, the indifference curve of the high-ability worker has a smaller slope. This property of preferences, known as the *single-crossing property*, plays an important role in the analysis of signaling models and in models of asymmetric information more generally. It arises here because the worker's marginal rate of substitution between wages and education at any given (w, e) pair is $(dw/de)_{\bar{u}} = c_e(e, \theta)$, which is decreasing in θ because $c_{e\theta}(e, \theta) < 0$.

We can also graph a function giving the equilibrium wage offer that results for each education level, which we denote by $w(e)$. Note that since in any PBE $w(e) = \mu(e)\theta_H + (1 - \mu(e))\theta_L$ for the equilibrium belief function $\mu(e)$, the equilibrium wage offer resulting from any choice of e must lie in the interval $[\theta_L, \theta_H]$. A possible wage offer function $w(e)$ is shown in Figure 13.C.3.

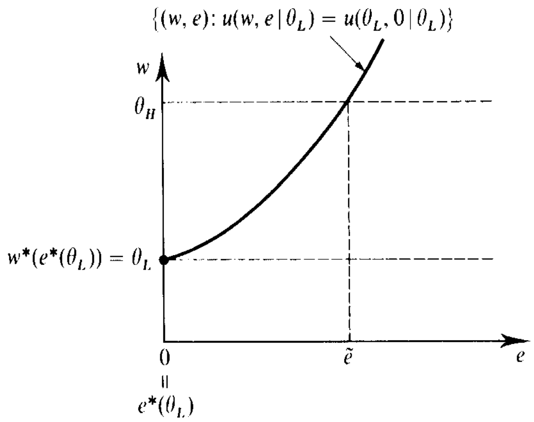
We are now ready to determine the equilibrium education choices for the two types of workers. It is useful to consider separately two different types of equilibria that might arise: *separating equilibria*, in which the two types of workers choose different education levels, and *pooling equilibria*, in which the two types choose the same education level.

Separating Equilibria

To analyze separating equilibria, let $e^*(\theta)$ be the worker's equilibrium education choice as a function of her type, and let $w^*(e)$ be the firms' equilibrium wage offer as a function of the worker's education level. We first establish two useful lemmas.

Lemma 13.C.1: In any separating perfect Bayesian equilibrium, $w^*(e^*(\theta_H)) = \theta_H$ and $w^*(e^*(\theta_L)) = \theta_L$; that is, each worker type receives a wage equal to her productivity level.

Proof: In any PBE, beliefs on the equilibrium path must be correctly derived from the equilibrium strategies using Bayes' rule. Here this implies that upon seeing education level $e^*(\theta_L)$, firms must assign probability one to the worker being type θ_L. Likewise, upon seeing education level $e^*(\theta_H)$, firms must assign probability one



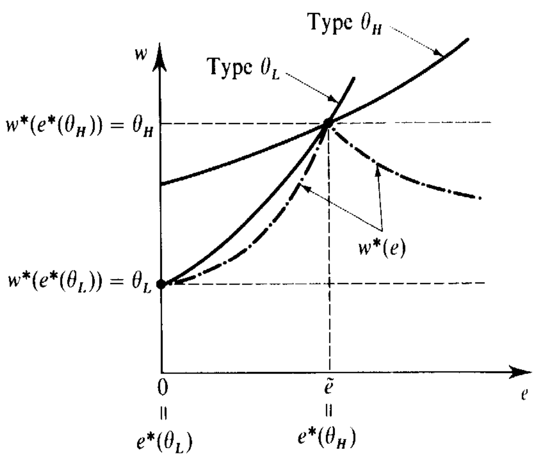


Figure 13.C.4 (left) Low-ability worker's outcome in a separating equilibrium.

Figure 13.C.5 (right) A separating equilibrium: Type is inferred from education level.

to the worker being type θ_H. The resulting wages are then exactly θ_L and θ_H, respectively. ■

Lemma 13.C.2: In any separating perfect Bayesian equilibrium, $e^*(\theta_L) = 0$; that is, a low-ability worker chooses to get no education.

Proof: Suppose not, that is, that when the worker is type θ_L, she chooses some strictly positive education level $\hat{e} > 0$. According to Lemma 13.C.1, by doing so, the worker receives a wage equal to θ_L. However, she would receive a wage of at least θ_L if she instead chose $e = 0$. Since choosing $e = 0$ would have save her the cost of education, she would be strictly better off by doing so, which is a contradiction to the assumption that $\hat{e} > 0$ is her equilibrium education level. ■

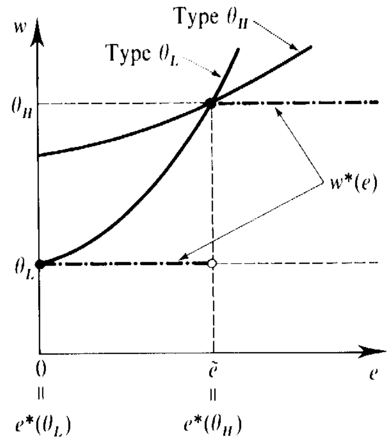
Lemma 13.C.2 implies that, in any separating equilibrium, type θ_L's indifference curve through her equilibrium level of education and wage must look as depicted in Figure 13.C.4.

Using Figure 13.C.4, we can construct a separating equilibrium as follows: Let $e^*(\theta_H) = \tilde{e}$, let $e^*(\theta_L) = 0$, and let the schedule $w^*(e)$ be as drawn in Figure 13.C.5. The firms' equilibrium beliefs following education choice e are $\mu^*(e) = (w^*(e) - \theta_L)/(\theta_H - \theta_L)$. Note that they satisfy $\mu^*(e) \in [0, 1]$ for all $e \geq 0$, since $w^*(e) \in [\theta_L, \theta_H]$.

To verify that this is indeed a PBE, note that we are completely free to let firms have any beliefs when e is neither 0 nor $\tilde{e}$. On the other hand, we must have $\mu(0) = 0$ and $\mu(\tilde{e}) = 1$. The wage offers drawn, which have $w^*(0) = \theta_L$ and $w^*(\tilde{e}) = \theta_H$, reflect exactly these beliefs.

What about the worker's strategy? It is not hard to see that, given the wage function $w^*(e)$, the worker is maximizing her utility by choosing $e = 0$ when she is type θ_L and by choosing $e = \tilde{e}$ when she is type θ_H. This can be seen in Figure 13.C.5 by noting that, for each type that she may be, the worker's indifference curve is at its highest-possible level along the schedule $w^*(e)$. Thus, strategies $[e^*(\theta), w^*(e)]$ and the associated beliefs $\mu(e)$ of the firms do in fact constitute a PBE.

Note that this is not the only PBE involving these education choices by the two types of workers. Because we have so much freedom to choose the firms' beliefs off the equilibrium path, many wage schedules can arise that support these education



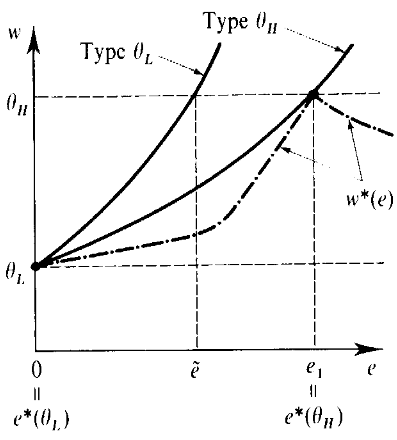


Figure 13.C.6 (left)
A separating equilibrium with the same education choices as in Figure 13.C.5 but different off-equilibrium-path beliefs.

Figure 13.C.7 (right)
A separating equilibrium with an education choice $e^*(\theta_H) > \tilde{e}$ by high-ability workers.

choices. Figure 13.C.6 depicts another one; in this PBE, firms believe that the worker is certain to be of high quality if $e \geq \tilde{e}$ and is certain to be of low quality if $e < \tilde{e}$. The resulting wage schedule has $w^*(e) = \theta_H$ if $e \geq \tilde{e}$ and $w^*(e) = \theta_L$ if $e < \tilde{e}$.

In these separating equilibria, high-ability workers are willing to get otherwise useless education simply because it allows them to distinguish themselves from low-ability workers and receive higher wages. The fundamental reason that education can serve as a signal here is that the marginal cost of education depends on a worker's type. Because the marginal cost of education is higher for a low-ability worker [since $c_{e\theta}(e, \theta) < 0$], a type θ_H worker may find it worthwhile to get some positive level of education $e' > 0$ to raise her wage by some amount $\Delta w > 0$, whereas a type θ_L worker may be unwilling to get this same level of education in return for the same wage increase. As a result, firms can reasonably come to regard education level as a signal of worker quality.

The education level for the high-ability type observed above is not the only one that can arise in a separating equilibrium in this model. Indeed, many education levels for the high-ability type are possible. In particular, any education level between $\tilde{e}$ and e_1 in Figure 13.C.7 can be the equilibrium education level of the high-ability workers. A wage schedule that supports education level $e^*(\theta_H) = e_1$ is depicted in the figure. Note that the education level of the high-ability worker cannot be below $\tilde{e}$ in a separating equilibrium because, if it were, the low-ability worker would deviate and pretend to be of high ability by choosing the high-ability education level. On the other hand, the education level of the high-ability worker cannot be above e_1 because, if it were, the high-ability worker would prefer to get no education, even if this resulted in her being thought to be of low ability.

Note that these various separating equilibria can be Pareto ranked. In all of them, firms earn zero profits, and a low-ability worker's utility is θ_L. However, a high-ability worker does strictly better in equilibria in which she gets a lower level of education. Thus, separating equilibria in which the high-ability worker gets education level $\tilde{e}$ (e.g., the equilibria depicted in Figures 13.C.5 and 13.C.6) Pareto dominate all the others. The Pareto-dominated equilibria are sustained because of the high-ability worker's fear that if she chooses a lower level of education than that prescribed in the equilibrium firms will believe that she is not a high-ability worker. These beliefs can be maintained because in equilibrium they are never disconfirmed.

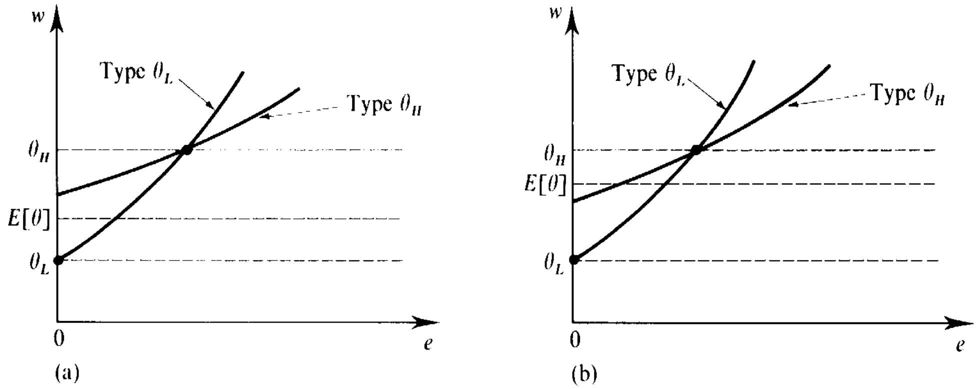


Figure 13.C.8
Separating equilibria may be Pareto dominated by the no-signaling outcome. (a) A separating equilibrium that is not Pareto dominated by the no-signaling outcome. (b) A separating equilibrium that is Pareto dominated by the no-signaling outcome.

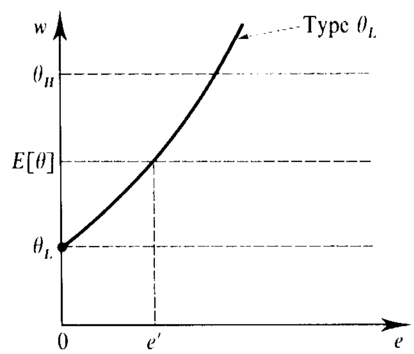
It is of interest to compare welfare in these equilibria with that arising when worker types are unobservable but no opportunity for signaling is available. When education is not available as a signal (so workers also incur no education costs), we are back in the situation studied in Section 13.B. In both cases, firms earn expected profits of zero. However, low-ability workers are strictly worse off when signaling is possible. In both cases they incur no education costs, but when signaling is possible they receive a wage of θ_L rather than $E(\theta)$.

What about high-ability workers? The somewhat surprising answer is that high-ability workers may be either better or worse off when signaling is possible. In Figure 13.C.8(a), the high-ability workers are better off because of the increase in their wages arising through signaling. However, in Figure 13.C.8(b), even though high-ability workers seek to take advantage of the signaling mechanism to distinguish themselves, they are *worse* off than when signaling is impossible! Although this may seem paradoxical (if high-ability workers choose to signal, how can they be worse off?), its cause lies in the fact that in a separating signaling equilibrium firms' expectations are such that the wage–education outcome from the no-signaling situation, $(w, e) = (E[\theta], 0)$, is no longer available to the high-ability workers; if they get no education in the separating signaling equilibrium, they are thought to be of low ability and offered a wage of θ_L. Thus, they can be worse off when signaling is possible, even though they are choosing to signal.

Note that because the set of separating equilibria is completely unaffected by the fraction λ of high-ability workers, as this fraction grows it becomes more likely that the high-ability workers are made worse off by the possibility of signaling [compare Figures 13.C.8(a) and 13.C.8(b)]. In fact, as this fraction gets close to 1, nearly every worker is getting costly education just to avoid being thought to be one of the handful of bad workers!

Pooling Equilibria

Consider now pooling equilibria, in which the two types of workers choose the same level of education, $e^*(\theta_L) = e^*(\theta_H) = e^*$. Since the firms' beliefs must be correctly derived from the equilibrium strategies and Bayes' rule when possible, their beliefs when they see education level e^* must assign probability λ to the worker being type θ_H. Thus, in any pooling equilibrium, we must have $w^*(e^*) = \lambda\theta_H + (1 - \lambda)\theta_L = E[\theta]$.



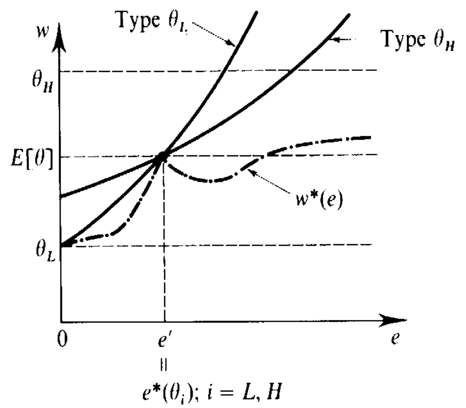


Figure 13.C.9 (left)
The highest-possible education level in a pooling equilibrium.

Figure 13.C.10 (right)
A pooling equilibrium.

The only remaining issue therefore concerns what levels of education can arise in a pooling equilibrium. It turns out that any education level between 0 and the level e' depicted in Figure 13.C.9 can be sustained.

Figure 13.C.10 shows an equilibrium supporting education level e'. Given the wage schedule depicted, each type of worker maximizes her payoff by choosing education level e'. This wage schedule is consistent with Bayesian updating on the equilibrium path because it gives a wage offer of $E[\theta]$ when education level e' is observed.

Education levels between 0 and e' can be supported in a similar manner. Education levels greater than e' cannot be sustained because a low-ability worker would rather set $e = 0$ than $e > e'$ even if this results in a wage payment of θ_L. Note that a pooling equilibrium in which both types of worker get no education Pareto dominates any pooling equilibrium with a positive education level. Once again, the Pareto-dominated pooling equilibria are sustained by the worker's fear that a deviation will lead firms to have an unfavorable impression of her ability. Note also that a pooling equilibrium in which both types of worker obtain no education results in exactly the same outcome as that which arises in the absence of an ability to signal. Thus, pooling equilibria are (weakly) Pareto dominated by the no-signaling outcome.

Multiple Equilibria and Equilibrium Refinement

The multiplicity of equilibria observed here is somewhat disconcerting. As we have seen, we can have separating equilibria in which firms learn the worker's type, but we can also have pooling equilibria where they do not; and within each type of equilibrium, many different equilibrium levels of education can arise. In large part, this multiplicity stems from the great freedom that we have to choose beliefs off the equilibrium path. Recently, a great deal of research has investigated the implications of putting "reasonable" restrictions on such beliefs along the lines we discussed in Section 9.D.

To see a simple example of this kind of reasoning, consider the separating equilibrium depicted in Figure 13.C.7. To sustain e_1 as the equilibrium education level of high-ability workers, firms must believe that any worker with an education level below e_1 has a positive probability of being of type θ_L. But consider any education level $\hat{e} \in (\tilde{e}, e_1)$. A type θ_L worker could never be made better off choosing such an education level than she is getting education level $e = 0$ *regardless of what* *firms believe about her as a result.* Hence, any belief by firms upon seeing education level $\hat{e} > \tilde{e}$ other than $\mu(\hat{e}) = 1$ seems unreasonable. But if this is so, then we must have $w(\hat{e}) = \theta_H$, and so the high-ability worker would deviate to $\hat{e}$. In fact, by this logic, the only education level that can be chosen by type θ_H workers in a separating equilibrium involving reasonable beliefs is $\tilde{e}$.

In Appendix A we discuss in greater detail the use of these types of reasonable-beliefs refinements. One refinement proposed by Cho and Kreps (1987), known as the *intuitive criterion*, extends the idea discussed in the previous paragraph to rule out not only the dominated separating equilibria but also all pooling equilibria. Thus, if we accept the Cho and Kreps (1987) argument, we predict a *unique* outcome to this two-type signaling game: the best separating equilibrium outcome, which is shown in Figures 13.C.5 and 13.C.6.

Second-Best Market Intervention

In contrast with the market outcome predicted by the game-theoretic model studied in Section 13.B (the highest-wage competitive equilibrium), in the presence of signaling a central authority who cannot observe worker types may be able to achieve a Pareto improvement relative to the market outcome. To see this in the simplest manner, suppose that the Cho and Kreps (1987) argument predicting the best separating equilibrium outcome is correct. We have already seen that the best separating equilibrium can be Pareto dominated by the outcome that arises when signaling is impossible. When it is, a Pareto improvement can be achieved simply by banning the signaling activity.

In fact, it may be possible to achieve a Pareto improvement even when the no-signaling outcome does not Pareto dominate the best separating equilibrium. To see how, consider Figure 13.C.11. In the figure, the best separating equilibrium has low-ability workers at point $(\theta_L, 0)$ and high-ability workers at point $(\theta_H, \tilde{e})$. Note that the high-ability workers would be worse off if signaling were banned, since the point $(E[\theta], 0)$ gives them less than their equilibrium level of utility. Nevertheless, note that if we gave the low- and high-ability workers outcomes of $(\hat{w}_L, 0)$ and $(\hat{w}_H, \hat{e}_H)$, respectively, both types would be better off. The central authority can achieve this outcome by mandating that workers with education levels below $\hat{e}_H$ receive a wage of $\hat{w}_L$ and that workers with education levels of at least $\hat{e}_H$ receive a

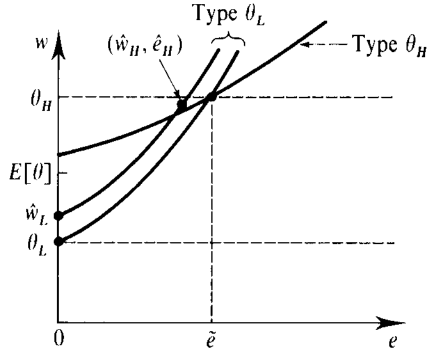


Figure 13.C.11
Achieving a Pareto improvement through cross-subsidization.

wage of $\hat{w}_H$. If so, low-ability workers would choose $e = 0$ and high-ability workers would choose $e = \hat{e}_H$. This alternative outcome involves firms incurring losses on low-ability workers and making profits on high-ability workers. However, as long as the firms break even on *average*, they are no worse off than before and a Pareto improvement has been achieved. The key to this Pareto improvement is that the central authority introduces *cross-subsidization*, where high-ability workers are paid less than their productivity level while low-ability workers are paid more than theirs, an outcome that cannot occur in a separating signaling equilibrium. (Note that the outcome when signaling is banned is an extreme case of cross-subsidization.)

Exercise 13.C.3: In the signaling model discussed in Section 13.C with $r(\theta_H) = r(\theta_L) = 0$, construct an example in which a central authority who does not observe worker types can achieve a Pareto improvement over the best separating equilibrium through a policy that involves cross-subsidization, but cannot achieve a Pareto improvement by simply banning the signaling activity. [*Hint*: Consider first a case with linear indifference curves.]

The case with $r(\theta_H) = r(\theta_L) = 0$ studied above, in which the market outcome in the absence of signaling is Pareto optimal, illustrates how the use of costly signaling can reduce welfare. Yet, when the market outcome in the absence of signaling is not efficient, signaling's ability to reveal information about worker types may instead create a Pareto improvement by leading to a more efficient allocation of labor. To see this point, suppose that we have $r = r(\theta_L) = r(\theta_H)$, with $\theta_L < r < \theta_H$ and $E[\theta] < r$. In this case, the equilibrium outcome without signaling has no workers employed. In contrast, any Pareto efficient outcome must have the high-ability workers employed by firms.

We now study the equilibrium outcome when signaling is possible. Consider, first, the wage and employment outcome that results after educational choice e by the worker. Following the worker's choice of educational level e, equilibrium behavior involves a wage of $w^*(e) = \mu(e)\theta_H + (1 - \mu(e))\theta_L$. If $w^*(e) \geq r$, then both types of workers would accept employment; if $w^*(e) < r$, then neither type would do so.

We now determine the equilibrium education choices of the two types of workers. Note first that any pooling equilibrium must have both types choosing $e = 0$ and neither type accepting employment. To see this, suppose that both types are choosing education level $\hat{e}$. Then $\mu(\hat{e}) = \lambda$ and $w^*(e) = E[\theta] < r$, and so neither type accepts employment. Hence, if $\hat{e} > 0$, both types would be better off choosing $e = 0$ instead. Thus, only an education level of zero is possible in a pooling equilibrium. In this zero education pooling equilibrium, the outcome is identical to the equilibrium outcome arising in the absence of the opportunity to signal.

The set of separating equilibria, on the other hand, is illustrated in Figure 13.C.12. In any separating equilibrium, a low-ability worker sets $e = 0$, is offered a wage of θ_L, and chooses to work at home, thereby achieving a utility of r. High-ability workers, on the other hand, select an education level in the interval $[\hat{e}, e_2]$ depicted in the figure, are offered a wage of θ_H, and accept employment. Note that no separating equilibrium can have $e^*(\theta_H) < \hat{e}$, since then low-ability workers would deviate and set $e = e^*(\theta_H)$; also, no separating equilibrium can have $e^*(\theta_H) > e_2$, since high-ability workers would then be better off setting $e = 0$ and working at home.

Note that in all these equilibria, both pooling and separating, the high-ability workers are weakly better off compared with the equilibrium arising without signaling opportunities and are strictly better off in separating equilibria with $e^*(\theta_H) < e_2$. Moreover, both the low-ability workers and the firms are equally well off. Thus, in the case with $\theta_L < r < \theta_H$ and $E[\theta] < r$, any pooling or separating signaling equilibrium weakly *Pareto dominates* the outcome arising

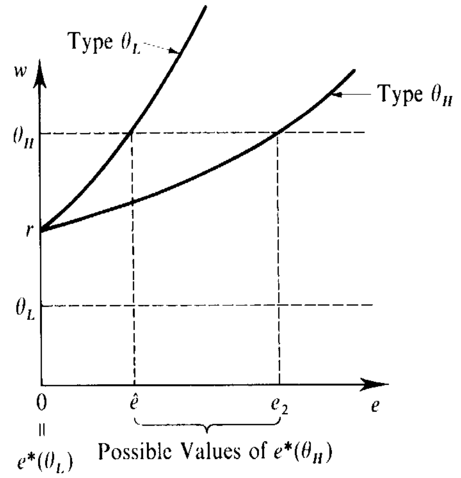


Figure 13.C.12
Separating equilibria when $r(\theta_L) = r(\theta_H) = r \in (\theta_L, \theta_H)$.

in the absence of signaling, and this Pareto dominance is *strict* for (essentially) all separating equilibria.

13.D Screening

In Section 13.C, we considered how signaling may develop in the marketplace as a response to the problem of asymmetric information about a good to be traded. There, individuals on the *more informed* side of the market (workers) chose their level of education in an attempt to signal information about their abilities to uninformed parties (the firms). In this section, we consider an alternative market response to the problem of unobservable worker productivity in which the *uninformed* parties take steps to try to distinguish, or *screen*, the various types of individuals on the other side of the market.[21] This possibility was first studied by Rothschild and Stiglitz (1976) and Wilson (1977) in the context of insurance markets (see Exercise 13.D.2).

As in Section 13.C, we focus on the case in which there are two types of workers, θ_L and θ_H, with $\theta_H > \theta_L > 0$ and where the fraction of workers who are of type θ_H is $\lambda \in (0, 1)$. In addition, workers earn nothing if they do not accept employment in a firm [in the notation used in Section 13.B, $r(\theta_L) = r(\theta_H) = 0$]. However, we now suppose that jobs may differ in the "task level" required of the worker. For example, jobs could differ in the number of hours per week that the worker is required to work. Or the task level might represent the speed at which a production line is run in a factory.

To make matters particularly simple, and to make the model parallel that in Section 13.C, we suppose that higher task levels add *nothing* to the output of the worker; rather, their *only* effect is to lower the utility of the worker.[22] The output of a type θ worker is therefore θ regardless of the worker's task level.

21. The setting analyzed here is one of *competitive screening* of workers, since we assume that there are several competing firms. See Section 14.C for a discussion of the *monopolistic screening* case, where a single firm screens workers.

22. As was true in the case of educational signaling, the assumption that higher task levels do not raise productivity is made purely for expositional purposes. Exercise 13.D.1 considers the case in which the firms' profits are increasing in the task level.

We assume that the utility of a type θ worker who receives wage w and faces task level $t \geq 0$ is

$$u(w, t \mid \theta) = w - c(t, \theta),$$

where $c(t, \theta)$ has all the properties assumed of the function $c(e, \theta)$ in Section 13.C. In particular, $c(0, \theta) = 0$, $c_t(t, \theta) > 0$, $c_{tt}(t, \theta) > 0$, $c_\theta(t, \theta) < 0$ for all $t > 0$, and $c_{t\theta}(t, \theta) < 0$. As will be clear shortly, the task level t serves to distinguish among types here in a manner that parallels the role of education in the signaling model discussed in Section 13.C.

Here we study the pure strategy subgame perfect Nash equilibria (SPNEs) of the following two-stage game:[23]

Stage 1: Two firms simultaneously announce sets of offered contracts. A contract is a pair (w, t). Each firm may announce any finite number of contracts.

Stage 2: Given the offers made by the firms, workers of each type choose whether to accept a contract and, if so, which one. For simplicity, we assume that if a worker is indifferent between two contracts, she always chooses the one with the lower task level and that she accepts employment if she is indifferent about doing so. If a worker's most preferred contract is offered by both firms, she accepts each firm's offer with probability $\frac{1}{2}$.

Thus, a firm can offer a variety of contracts; for example, it might have several production lines, each running at a different speed. Different types of workers may then end up choosing different contracts.[24]

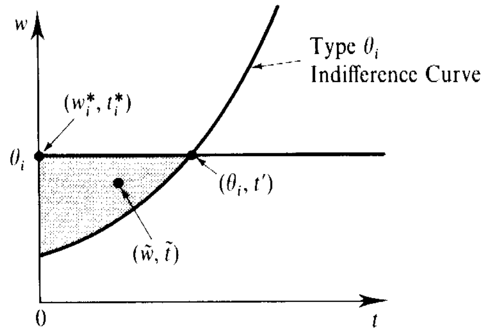
It is helpful to start by considering what the outcome of this game would be if worker types were *observable*. To address this case, we allow firms to condition their offer on a worker's type (so that a firm can offer a contract (w_L, t_L) solely to type θ_L workers and another contract (w_H, t_H) solely to type θ_H workers).

Proposition 13.D.1: In any SPNE of the screening game with observable worker types, a type θ_i worker accepts contract $(w_i^*, t_i^*) = (\theta_i, 0)$, and firms earn zero profits.

Proof: We first argue that any contract (w_i^*, t_i^*) that workers of type θ_i accept in equilibrium must produce exactly zero profits; that is, it must involve a wage $w_i^* = \theta_i$. To see this, note that if $w_i^* > \theta_i$, then some firm is making a loss offering this contract and it would do better by not offering any contract to type θ_i workers. Suppose, on the other hand, that $w_i^* < \theta_i$, and let $\Pi > 0$ be the aggregate profits earned by the two firms on type θ_i workers. One of the two firms must be earning no more than $\Pi/2$ from these workers. If it deviates by offering a contract $(w_i^* + \varepsilon, t_i^*)$ for any

23. For this game, the set of subgame perfect Nash equilibria is identical to the sets of strategy profiles in weak perfect Bayesian equilibria or sequential equilibria.

24. The models in the original Rothschild and Stiglitz (1976) and Wilson (1977) analyses differ from our model in two respects. First, firms in those papers were restricted to offering only a single contract. This could make sense in the production line interpretation, for example, if each firm had only a single production line. Second, those authors allowed for "free entry," so that an additional firm could always enter if a profitable contracting opportunity existed. In fact, making these two changes has little effect on our conclusions. The only difference is in the precise conditions under which an equilibrium exists. (For more on this, see Exercise 13.D.4.)



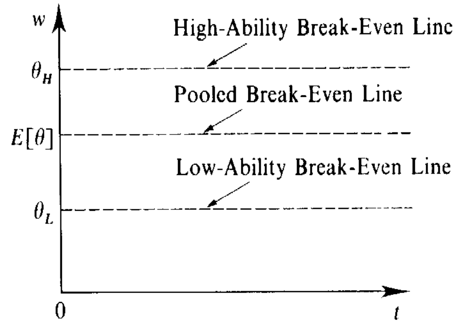


Figure 13.D.1 (left) The equilibrium contract (w_i^*, t_i^*) for type θ_i with perfect observability.

Figure 13.D.2 (right) Break-even lines.

$\varepsilon > 0$, it will attract all type θ_i workers. Since ε can be made arbitrarily small, its profits from type θ_i workers can be made arbitrarily close to Π, and so this deviation will increase its profits. Thus, we must have $w_i^* = \theta_i$.

Now suppose that $(w_i^*, t_i^*) = (\theta_i, t')$ for some $t' > 0$. Then, as shown in Figure 13.D.1 (where the wage is measured on the vertical axis and the task level is measured on the horizontal axis), either firm could deviate and earn strictly positive profits by offering a contract in the shaded area of the figure, such as $(\tilde{w}, \tilde{t})$. The only contract at which there are no profitable deviations is $(w_i^*, t_i^*) = (\theta_i, 0)$, the contract that maximizes a type θ_i worker's utility subject to the constraint that the firms offering the contract break even. ■

We now turn to the situation in which worker types are *not observable*. In this case, each contract offered by a firm may in principle be accepted by either type of worker. We can note immediately that the complete information outcome identified in Proposition 13.D.1 cannot arise when worker types are unobservable: Because every low-ability worker prefers the high-ability contract $(\theta_H, 0)$ to contract $(\theta_L, 0)$, if these were the two contracts being offered by the firms then *all* workers would accept contract $(\theta_H, 0)$ and the firms would end up losing money.

To determine the equilibrium outcome with unobservable worker types, it is useful to begin by drawing three break-even lines: the zero-profit lines for productivity levels θ_L, $E[\theta]$, and θ_H, respectively. These three break-even lines are depicted by the dashed lines in Figure 13.D.2. The middle line represents the break-even line for a contract that attracts both types of workers, and we therefore refer to it as the *pooled* break-even line.

As in Section 13.C, we can in principle have two types of (pure strategy) equilibria: *separating* equilibria, in which the two types of workers accept different contracts, and *pooling* equilibria, in which both types of workers sign the same contract. (It can be shown that in any equilibrium both types of workers will accept some contract; we assume that this is so in the discussion that follows.) We proceed with a series of lemmas. Lemma 13.D.1 applies to both pooling and separating equilibria.

Lemma 13.D.1: In any equilibrium, whether pooling or separating, both firms must earn zero profits.

Proof: Let (w_L, t_L) and (w_H, t_H) be the contracts chosen by the low- and high-ability workers, respectively (these could be the same contract), and suppose that the two firms' aggregate profits are $\Pi > 0$. Then one firm must be making no more than $\Pi/2$. Consider a deviation by this firm in which it offers contracts $(w_L + \varepsilon, t_L)$ and

$(w_H + \varepsilon, t_H)$ for $\varepsilon > 0$. Contract $(w_L + \varepsilon, t_L)$ will attract all type θ_L workers, and contract $(w_H + \varepsilon, t_H)$ will attract all type θ_H workers. [Note that since type θ_i initially prefers contract (w_i, t_i) to (w_j, t_j), we have $w_i - c(t_i, \theta_i) \geq w_j - c(t_j, \theta_i)$, and so $(w_i + \varepsilon) - c(t_i, \theta_i) \geq (w_j + \varepsilon) - c(t_j, \theta_i)$.] Since ε can be chosen to be arbitrarily small, this deviation yields this firm profits arbitrarily close to Π, and so the firm has a profitable deviation. Thus, we must have $\Pi \leq 0$. Because no firm can incur a loss in any equilibrium (it could always earn zero by offering no contracts), both firms must in fact earn a profit of zero. ■

An important implication of Lemma 13.D.1 is that, in any equilibrium, no firm can have a deviation that allows it to earn strictly positive profits. We shall use this fact repeatedly in the discussion that follows. Using it, we immediately get the result given in Lemma 13.D.2 regarding pooling equilibria.

Lemma 13.D.2: No pooling equilibria exist.

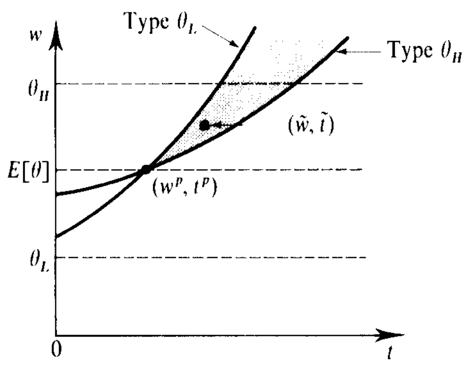
Proof: Suppose that there is a pooling equilibrium contract (w^P, t^P). By Lemma 13.D.1, it lies on the pooled break-even line, as shown in Figure 13.D.3. Suppose that firm j is offering contract (w^P, t^P). Then firm $k \neq j$ has a deviation that yields it a strictly positive profit: It offers a single contract $(\tilde{w}, \tilde{t})$ that lies somewhere in the shaded region in Figure 13.D.3 and has $\tilde{w} < \theta_H$. This contract attracts all the type θ_H workers and none of the type θ_L workers, who prefer (w^P, t^P) over $(\tilde{w}, \tilde{t})$. Moreover, since $\tilde{w} < \theta_H$, firm k makes strictly positive profits from this contract when the high-ability workers accept it. ■

We now consider the possibilities for separating equilibria. Lemma 13.D.3 shows that all contracts accepted in a separating equilibrium must yield zero profits.

Lemma 13.D.3: If (w_L, t_L) and (w_H, t_H) are the contracts signed by the low- and high-ability workers in a separating equilibrium, then both contracts yield zero profits; that is, $w_L = \theta_L$ and $w_H = \theta_H$.

Proof: Suppose first that $w_L < \theta_L$. Then either firm could earn strictly positive profits by instead offering only contract $(\tilde{w}_L, t_L)$, where $\theta_L > \tilde{w}_L > w_L$. All low-ability workers would accept this contract; moreover, the deviating firm earns strictly positive profits from any worker (of low or high ability) who accepts it. Since Lemma 13.D.1 implies that no such deviation can exist in an equilibrium, we must have $w_L \geq \theta_L$ in any separating equilibrium.

Suppose, instead, that $w_H < \theta_H$, as in Figure 13.D.4. If we have a separating



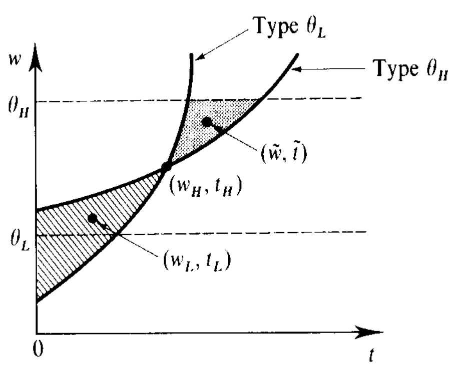


Figure 13.D.3 (left)
No pooling equilibria exist.

Figure 13.D.4 (right)
The high-ability contract in a separating equilibrium cannot have $w_H < \theta_H$.

equilibrium, then the type θ_L contract (w_L, t_L) must lie in the hatched region of the figure (by Lemma 13.D.1, it must also have $w_L > \theta_L$). To see this, note that since type θ_H workers choose contract (w_H, t_H), contract (w_L, t_L) must lie on or below the type θ_H indifference curve through (w_H, t_H), and since type θ_L workers choose (w_L, t_L) over (w_H, t_H), contract (w_L, t_L) must lie on or above the type θ_L indifference curve through (w_H, t_H). Suppose that firm j is offering the low-ability contract (w_L, t_L). Then firm $k \neq j$ could earn strictly positive profits by deviating and offering only a contract lying in the shaded region of the figure with a wage strictly less than θ_H, such as $(\tilde{w}, \tilde{t})$. This contract, which has $w_H < \theta_H$, will be accepted by all the type θ_H workers and by none of the type θ_L workers [since firm j will still be offering contract (w_L, t_L)]. So we must have $w_H \geq \theta_H$ in any separating equilibrium.

Since, by Lemma 13.D.1, firms break even in any equilibrium, we must in fact have $w_L = \theta_L$ and $w_H = \theta_H$. ■

Lemma 13.D.4 identifies the contract that must be accepted by low-ability workers in any separating equilibrium.

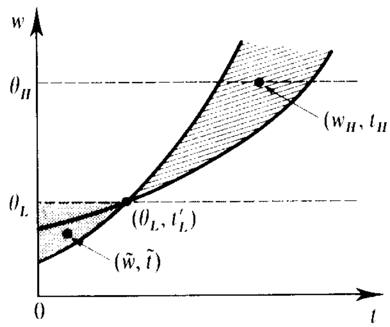
Lemma 13.D.4: In any separating equilibrium, the low-ability workers accept contract $(\theta_L, 0)$; that is, they receive the same contract as when no informational imperfections are present in the market.

Proof: By Lemma 13.D.3, $w_L = \theta_L$ in any separating equilibrium. Suppose that the low-ability workers' contract is instead some point (θ_L, t'_L) with $t'_L > 0$, as in Figure 13.D.5. (Although it is not important for the proof, the high-ability contract must then lie on the segment of the high-ability break-even line lying in the hatched region of the figure, as shown.) If so, then a firm can make strictly positive profits by offering only a contract lying in the shaded region of the figure, such as $(\tilde{w}, \tilde{t})$. All low-ability workers accept this contract, and the contract yields the firm strictly positive profits from any worker (of low or high ability) who accepts it. ■

We can now derive the high-ability workers' contract.

Lemma 13.D.5: In any separating equilibrium, the high-ability workers accept contract $(\theta_H, \hat{t}_H)$, where $\hat{t}_H$ satisfies $\theta_H - c(\hat{t}_H, \theta_L) = \theta_L - c(0, \theta_L)$.

Proof: Consider Figure 13.D.6. By Lemmas 13.D.3 and 13.D.4, we know that $(w_L, t_L) = (\theta_L, 0)$ and that $w_H = \theta_H$. In addition, if the type θ_L workers are willing to accept contract $(\theta_L, 0)$, t_H must be at least as large as the level $\hat{t}_H$ depicted in the



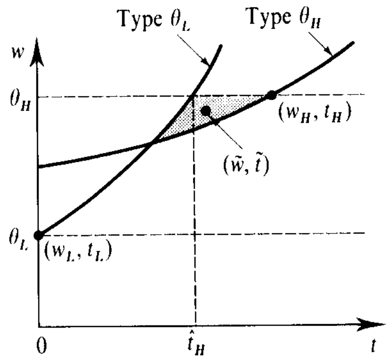


Figure 13.D.5 (left)
The low-ability workers must receive contract $(\theta_L, 0)$ in any separating equilibrium.

Figure 13.D.6 (right)
The high-ability workers must receive contract $(\theta_H, \hat{t}_H)$ in any separating equilibrium.

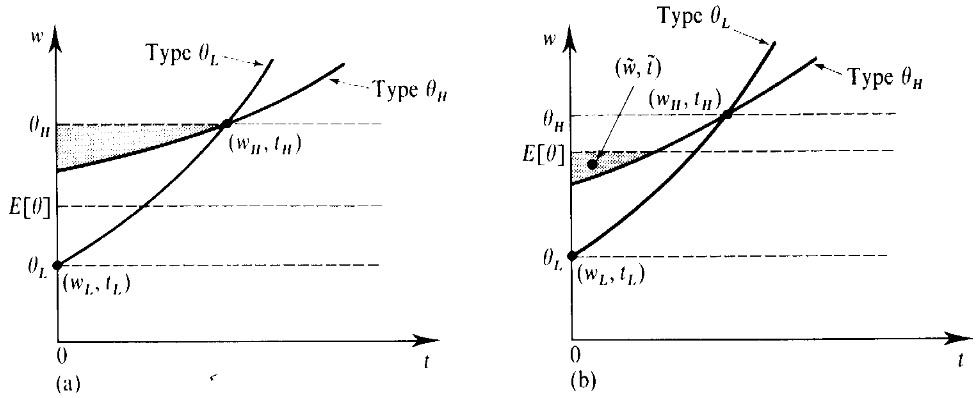


Figure 13.D.7
An equilibrium may not exist. (a) No pooling contract breaks the separating equilibrium. (b) The pooling contract $(\tilde{w}, \tilde{t})$ breaks the separating equilibrium.

figure. Note that low-ability workers are indifferent between contracts $(\theta_L, 0)$ and $(\theta_H, \hat{t}_H)$, and so $\theta_H - c(\hat{t}_H, \theta_L) = \theta_L - c(0, \theta_L)$. Suppose, then, that the high-ability contract (θ_H, t_H) has $t_H > \hat{t}_H$, as in the figure. Then either firm can earn a strictly positive profit by also offering, in addition to its current contracts, a contract lying in the shaded region of the figure with $w_H < \theta_H$, such as $(\tilde{w}, \tilde{t})$. This contract attracts all the high-ability workers and does not change the choice of the low-ability workers. Thus, in any separating equilibrium, the high-ability contract must be $(\theta_H, \hat{t}_H)$. ■

Proposition 13.D.2 summarizes the discussion so far.

Proposition 13.D.2: In any subgame perfect Nash equilibrium of the screening game, low-ability workers accept contract $(\theta_L, 0)$, and high-ability workers accept contract $(\theta_H, \hat{t}_H)$, where $\hat{t}_H$ satisfies $\theta_H - c(\hat{t}_H, \theta_L) = \theta_L - c(0, \theta_L)$.

Proposition 13.D.2 does not complete our analysis, however. Although we have established what any equilibrium must look like, we have not established that one exists. In fact, we now show that *one may not exist*.

Suppose that both firms are offering the two contracts identified in Proposition 13.D.2 and illustrated in Figure 13.D.7(a). Does either firm have an incentive to deviate? No firm can earn strictly positive profits by deviating in a manner that attracts either only high-ability or only low-ability workers (just try to find such a deviation). But what about a deviation that attracts *all* workers? Consider a deviation in which the deviating firm attracts all workers to a single pooling contract. In Figure 13.D.7(a), a contract can attract both types of workers if and only if it lies in the shaded region. There is no profitable deviation of this type if, as depicted in the figure, this shaded area lies completely above the pooled break-even line. However, when some of the shaded area lies strictly below the pooled break-even line, as in Figure 13.D.7(b), a profitable deviation to a pooling contract such as $(\tilde{w}, \tilde{t})$ exists. In this case, *no equilibrium exists*.

Even when no single pooling contract breaks the separating equilibrium, it is possible that a profitable deviation involving a pair of contracts may do so. For example, a firm can attract both types of workers by offering the contracts $(\tilde{w}_L, \tilde{t}_L)$ and $(\tilde{w}_H, \tilde{t}_H)$ depicted in Figure 13.D.8. When it does so, type θ_L workers accept contract $(\tilde{w}_L, \tilde{t}_L)$ and type θ_H workers accept $(\tilde{w}_H, \tilde{t}_H)$. If this pair of contracts yields the firm a positive profit, then this deviation breaks the separating contracts identified

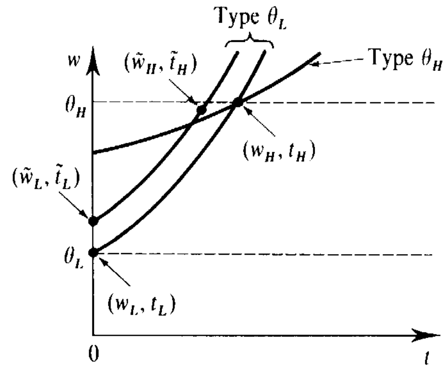


Figure 13.D.8
A profitable deviation using a pair of contracts may exist that breaks the separating equilibrium.

in Proposition 13.D.2 and no equilibrium exists. More generally, an equilibrium exists only if there is no such profitable deviation.

Welfare Properties of Screening Equilibria

Restricting attention to cases in which an equilibrium does exist, the screening equilibrium has welfare properties parallel to those of the signaling model's best separating equilibrium [with $r(\theta_L) = r(\theta_H) = 0$]. First, as in the earlier model, asymmetric information leads to Pareto inefficient outcomes. Here high-ability workers end up signing contracts that make them engage in completely unproductive and disutility-producing tasks merely to distinguish themselves from their less able counterparts. As in the signaling model, the low-ability workers are always worse off here when screening is possible than when it is not. One difference from the signaling model, however, is that in cases where an equilibrium exists, screening must make the high-ability workers better off; it is precisely in those cases where it would not that a move to a pooling contract breaks the separating equilibrium [see Figure 13.D.7(b)]. Indeed, when an equilibrium does exist, it is a constrained Pareto optimal outcome; if no firm has a deviation that can attract both types of workers and yield it a positive profit, then a central authority who is unable to observe worker types cannot achieve a Pareto improvement either.[25]

What can be said about the potential nonexistence of equilibrium in this model? Two paths have been followed in the literature. One approach is to establish existence of equilibria in the larger strategy space that allows for mixed strategies; on this, see Dasgupta and Maskin (1986). The other is to take the position that the lack of equilibria indicates that, in some important way, the model is incompletely specified. The aspect the literature has emphasized in this regard is the lack of any dynamic reactions to new contract offers [see Wilson (1977), Riley (1979), and Hellwig (1986)]. Wilson (1977), for example, uses a definition of equilibrium that captures the idea that firms are able to withdraw unprofitable contracts from the market. A set of contracts is a *Wilson equilibrium* if no firm has a profitable deviation that remains profitable once existing contracts that lose money after the deviation are withdrawn. This extra requirement may make deviations less attractive. In the deviation considered in Figure 13.D.3, for example, once contract $(\tilde{w}, \tilde{t})$ is introduced, the original contract (w^p, t^p) loses

25. Actually, there is a small gap: An equilibrium may exist when there is another pair of contracts that would give higher utility to both types of workers and that would yield the firm deviating to it exactly zero profits. In this case, the equilibrium is not a constrained Pareto optimum.

money. But if (w^p, t^p) is withdrawn as a result, then low-ability workers will accept $(\tilde{w}, \tilde{t})$ and this deviation ends up being unprofitable. Hellwig (1986) examines sequential equilibria and their refinements in a game that explicitly allows for such withdrawals.

By introducing such reactions, these papers establish the existence of pure strategy equilibria. Introducing reactions of this sort does not simply eliminate the nonexistence problem, however, but also yields somewhat different predictions regarding the characteristics of market equilibria and their welfare properties. For example, when firms can make multiple offers as we have allowed here, cross-subsidization can arise in Wilson equilibria. Indeed, Miyazaki (1977) shows that in the case in which multiple offers are possible, a Wilson equilibrium always exists and is necessarily a constrained Pareto optimum.

In the screening model examined above, we took the view that the uninformed firms made employment offers to the informed workers. Yet we could equally well imagine a model in which informed workers instead make contract offers to the firms. For example, each worker might propose a task level at which she is willing to work, and firms might then offer a wage for that task level. Note, however, that this alternative model exactly parallels the signaling model in Section 13.C and, as we have seen, yields quite different predictions. For example, the signaling model has numerous equilibria, but here we have at most a single equilibrium. This is somewhat disturbing. Given that our models are inevitably simplifications of actual market processes, if market outcomes are really very sensitive to issues such as this our models may provide us with little predictive ability.

One approach to this problem is offered by Maskin and Tirole (1992). They note that contracts like those we have allowed firms to offer in the screening model discussed in this section are still somewhat restricted. In particular, we could imagine a firm offering a worker a contract that involved an ex post (after signing) choice among a set of wage–task pairs (you will see more about contracts of this type in Section 14.C). Similarly, in considering the counterpart model in which workers make offers, we could allow a worker to propose such a contract. Maskin and Tirole (1992) show that with this enrichment of the allowed contracts (and a weak additional assumption) the sets of sequential equilibria of the two models coincide (there may be multiple equilibria in both cases).

APPENDIX A: REASONABLE-BELIEFS REFINEMENTS IN SIGNALING GAMES

In this appendix, we describe several commonly used reasonable-beliefs refinements of the perfect Bayesian and sequential equilibrium concepts for signaling games, and we apply them to the education signaling model discussed in Section 13.C. Excellent sources for further details and discussion are Cho and Kreps (1987) and Fudenberg and Tirole (1992).

Consider the following class of signaling games: There are I players plus nature. The first move of the game is nature's, who picks a "type" for player 1, $\theta \in \Theta = \{\theta_1, \ldots, \theta_N\}$. The probability of type θ is $f(\theta)$, and this is common knowledge among the players. However, only player 1 observes θ. The second move is player 1's, who picks an action a from set A after observing θ. Then, after seeing player 1's action choice (but not her type), each player $i = 2, \ldots, I$ simultaneously chooses an action s_i from set S_i. We define $S = S_2 \times \cdots \times S_I$. If player 1 is of type θ, her utility from choosing action a and having players $2, \ldots, I$ choose $s = (s_2, \ldots, s_I)$ is $u_1(a, s, \theta)$. Player $i \neq 1$ receives payoff $u_i(a, s, \theta)$ in this event. A perfect Bayesian equilibrium (PBE) in the sense used in Section 13.C is a profile of strategies $(a(\theta), s_2(a), \ldots, s_I(a))$, combined with a common belief function $\mu(\theta \mid a)$ for players $2, \ldots, I$ that assigns a probability $\mu(\theta \mid a)$ to type θ of player 1 conditional on observing action $a \in A$, such that

(i) Player 1's strategy is optimal given the strategies of players $2, \ldots, I$.
(ii) The belief function $\mu(\theta \mid a)$ is derived from player 1's strategy using Bayes' rule where possible.
(iii) The strategies of players $2, \ldots, I$ specify actions following each choice $a \in A$ that constitute a Nash equilibrium of the simultaneous-move game in which the probability that player 1 is of type θ is $\mu(\theta \mid a)$ for all $\theta \in \Theta$.

In the context of the model under study here, this notion of a PBE is equivalent to the sequential equilibrium notion.

The education signaling model in Section 13.C falls into this category of signaling games if we do not explicitly model the worker's choice between the firms' offers and instead simply incorporate into the payoff functions the implications of her optimal choice (she chooses from among the firms offering the highest wage if this wage is positive and refuses both firms' offers otherwise). In that model, $I = 3$, $\Theta = \{\theta_L, \theta_H\}$, the set $A = \{e: e \geq 0\}$ contains the possible education choices of the worker, and the set $S_i = \{w: w \in \mathbb{R}\}$ contains the possible wage offers by firm i.

Domination-Based Refinements of Beliefs

The simplest reasonable-belief refinement of the PBE notion arises from the idea (discussed in Section 9.D) that reasonable beliefs should not assign positive probability to a player taking an action that is strictly dominated for her. In a signaling game, this problem can arise when players $2, \ldots, I$ (the firms in the education signaling model) assign a probability $\mu(\theta \mid a) > 0$ to player 1 (the worker) being of type θ after observing action a, even though action a is a strictly dominated choice for player 1 when she is of type θ.

Formally, we say that action $a \in A$ is a strictly dominated choice for type θ if there is an action $a' \in A$ such that

$$\underset{s' \in S}{\text{Min}}\, u_1(a', s', \theta) > \underset{s \in S}{\text{Max}}\, u_1(a, s, \theta).^{26} \qquad (13.\text{AA}.1)$$

For each action $a \in A$, it is useful to define the set

$$\Theta(a) = \{\theta\text{: there is no } a' \in A \text{ satisfying (13.AA.1)}\}.$$

This is the set of types of player 1 for whom action a is not a strictly dominated choice. We can then say that a PBE has reasonable beliefs if, for all $a \in A$ with $\Theta(a) \neq \varnothing$,

$$\mu(\theta \mid a) > 0 \quad \text{only if} \quad \theta \in \Theta(a)$$

and we consider a PBE to be a sensible prediction only if it has reasonable beliefs.[27]

26. Note that a strategy $a(\theta)$ is strictly dominated for player 1 if and only if it involves play of a strictly dominated action for some type θ.

27. Doing this is equivalent to first eliminating each type θ's dominated actions from the game and then identifying the PBEs of this simplified game.

Unfortunately, in the education signaling model discussed in Section 13.C, this refinement does not narrow down our predictions at all. The set $\Theta(e)$ equals $\{\theta_L, \theta_H\}$ for all education levels e because either worker type will find e to be her optimal choice if the wage offered in response to e is sufficiently in excess of the wage offered at other education levels. Thus, no beliefs are ruled out, and all PBEs of the signaling game pass this test. If we want to narrow down our predictions for this model, we need to go beyond the use of refinements based only on notions of strict dominance.[28]

Recall the argument we made in Section 13.C for eliminating all separating equilibria but the best one. We argued that since, in Figure 13.C.7, a worker of type θ_L would be better off choosing $e = 0$ than she would choosing an education level above $\tilde{e}$ *for any beliefs and resulting equilibrium wage that might follow these two education levels*, no reasonable belief should assign a positive probability to a worker of type θ_L choosing any $e > \tilde{e}$. This is close to an argument that education levels $e > \tilde{e}$ are dominated choices for a type θ_L worker, but with the critical difference reflected in the italicized phrase: Only *equilibrium* responses of the firms are considered, rather than all conceivable responses. That is, we take a backward-induction-like view that the worker should only concern herself with possible equilibrium reactions to her education choices.

To be more formal about this idea, for any nonempty set $\hat{\Theta} \subset \Theta$, let $S^*(\hat{\Theta}, a) \subset S_2 \times \cdots \times S_I$ denote the set of possible equilibrium responses that can arise after action a is observed for *some* beliefs satisfying the property that $\mu(\theta \mid a) > 0$ only if $\theta \in \hat{\Theta}$. The set $S^*(\hat{\Theta}, a)$ contains the set of equilibrium responses by players $2, \ldots, I$ that can follow action choice a for some beliefs that assign positive probability only to types in $\hat{\Theta}$. When $\hat{\Theta} = \Theta$, the set of all conceivable types of player 1, this construction allows for all possible beliefs.[29] We can now say that action $a \in A$ is strictly dominated for type θ in this stronger sense if there exists an action a' with

$$\underset{s' \in S^*(\Theta, a')}{\text{Min}} \; u_1(a', s', \theta) > \underset{s \in S^*(\Theta, a)}{\text{Max}} \; u_1(a, s, \theta). \tag{13.AA.2}$$

Using this stronger notion of dominance, we can define the set

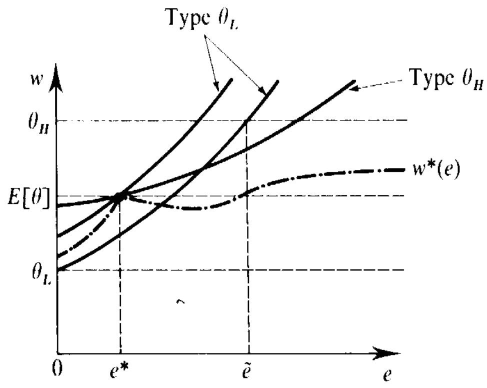
$$\Theta^*(a) = \{\theta\text{: there is no } a' \in A \text{ satisfying (13.AA.2)}\},$$

containing those types of player 1 for whom action a is not strictly dominated in the sense of (13.AA.2). We can now say that a PBE has reasonable beliefs if for all $a \in A$ with $\Theta^*(a) \neq \varnothing$, $\mu(a, \theta) > 0$ only if $\theta \in \Theta^*(a)$.

Using this reasonable-beliefs refinement significantly reduces the set of possible outcomes in the educational signaling model, sometimes even to a unique prediction. In that model, $S^*(\Theta, e) = [\theta_L, \theta_H]$ for all education choices e because, for any belief $\mu \in [0, 1]$, the resulting Nash equilibrium wage must lie between θ_L and θ_H. As a

28. We could, in principle, go further with this identification of strictly dominated strategies for player 1 by also eliminating any strictly dominated strategies for players $2, \ldots, I$, then looking to see whether we have any more strictly dominated actions for any of player 1's types, and so on. However, in the educational signaling model, this does not help us because the firms have no strictly dominated strategies.

29. Note that when there is only one player responding (so $I = 2$), the set $S^*(\Theta, a)$ is exactly the set of responses that are not strictly dominated for player 2 conditional on following action a. Note also that in this case a strategy $s_2(a)$ is weakly dominated for player 2 if, for any $a \in A$, it involves play of some $s \notin S^*(\Theta, a)$.



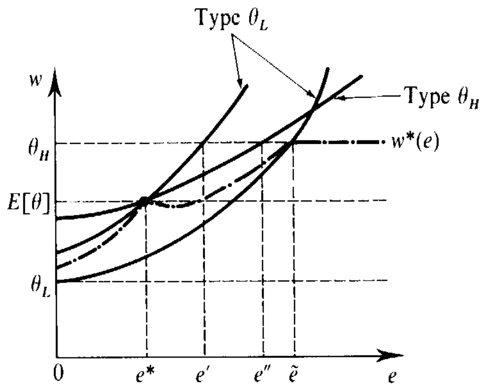


Figure 13.AA.1 (left)
A pooling equilibrium that is eliminated using the dominance test in (13.AA.2).

Figure 13.AA.2 (right)
A pooling equilibrium that is eliminated using the dominance test in (13.AA.3).

consequence, an education choice in excess of $\tilde{e}$ in Figure 13.C.7 is dominated for a type θ_L worker according to the test in (13.AA.2) by the education choice $e = 0$. Hence, in any PBE with reasonable beliefs, $\mu(\theta_H \mid e) = 1$ for all $e > \tilde{e}$. But if this is so, then no separating equilibrium with $e^*(\theta_H) > \tilde{e}$ can survive because, as we argued in Section 13.C, the high-ability worker will do better by deviating to an education level slightly in excess of $\tilde{e}$. Furthermore, we can also eliminate any pooling equilibrium in which the equilibrium outcome is worse for a high-ability worker than outcome $(\theta_H, \tilde{e})$, such as in the equilibrium depicted in Figure 13.AA.1, since any such equilibrium must involve unreasonable beliefs: If $\mu(\theta_H \mid e) = 1$ for all $e > \tilde{e}$, then a type θ_H worker could do better deviating to an education level just above $\tilde{e}$ where she would receive a wage of θ_H. In fact, when the high-ability worker prefers outcome $(\theta_H, \tilde{e})$ to $(E[\theta], 0)$, this argument rules out all pooling equilibria, and so we get the unique prediction of the best separating equilibrium.

Equilibrium Domination and the Intuitive Criterion

We now consider a further strengthening of the notion of dominance, known as *equilibrium dominance*. This leads to a refinement known as the *intuitive criterion* [Cho and Kreps (1987)] that always gives us the unique prediction of the best separating equilibrium in the two-type education signaling model studied in Section 13.C.

The idea behind this refinement can be seen by considering the pooling equilibrium of the education signaling model that is shown in Figure 13.AA.2, an equilibrium that is not eliminated by our previous refinements. Note that, as illustrated in the figure, to support education choice e^* as a pooling equilibrium outcome we must have beliefs for the firms satisfying $\mu(\theta_H \mid e) < 1$ for all $e \in (e', e'')$. Indeed, if $\mu(\theta_H \mid e) = 1$ at any such education level, then the wage offered would be θ_H and the type θ_H worker would find it optimal to deviate.

Suppose, however, that a firm is confronted with a deviation to some education level $\hat{e} \in (e', e'')$ when it was expecting the equilibrium level of education e^* to be chosen. It might reason as follows: "Either type of worker could be sure of getting outcome $(w, e) = (E[\theta], e^*)$ by choosing the equilibrium education level e^*. But a low-ability worker would be worse off deviating to education level e' *regardless* of what beliefs firms have after this choice, while a high-ability worker might be made better off by doing this. Thus, this must not be a low-ability worker." In this case, the choice of e' by the low-ability worker is dominated by her *equilibrium* payoff.

To formalize this idea in terms of our general specification, denote the equilibrium payoff to type θ in PBE $(a^*(\theta), s^*(a), \mu)$ by $u_1^*(\theta) = u_1(a^*(\theta), s^*(a^*(\theta)), \theta)$. We then say that action a is *equilibrium dominated* for type θ in PBE $(a^*(\theta), s^*(a), \mu)$ if

$$u_1^*(\theta) > \underset{s \in S^*(\Theta, a)}{\text{Max}} \; u_1(a, s, \theta). \tag{13.AA.3}$$

Using this notion of dominance, define for each $a \in A$ the set $\Theta^{**}(a) = \{\theta$: condition (13.AA.3) does not hold$\}$. We can now say that a PBE has reasonable beliefs if for all actions a with $\Theta^{**}(a) \neq \varnothing$, $\mu(\theta | a) > 0$ only if $\theta \in \Theta^{**}(a)$, and we can restrict attention to those PBEs that have reasonable beliefs.

Note that any action a that is dominated in the sense of (13.AA.2) for type θ must also be equilibrium dominated for this type because $u_1^*(\theta) = u_1^*(a^*(\theta), s^*(a^*(\theta)), \theta) > \text{Min}_{s' \in S^*(\Theta, a')} u_1(a', s', \theta)$ by the definition of a PBE. Thus, this equilibrium dominance-based procedure must rule out all the PBEs that were ruled out by our earlier procedure and may rule out more.

Consider the use of this refinement in the education signaling model of Section 13.C. Since it is stronger than the refinement based on (13.AA.2), this refinement also eliminates all but the best separating equilibrium. However, unlike our earlier dominance-based refinements, the equilibrium dominance-based refinement also eliminates *all* pooling equilibria. For example, in the pooling equilibrium depicted in Figure 13.AA.2, any education choice $\hat{e} \in (e', e'')$ is equilibrium dominated for the low-ability worker. Moreover, once the firms' beliefs following this education choice are restricted to assigning probability 1 to the worker being type θ_H, the high-ability worker wishes to deviate to this education level. Thus, we get a unique prediction for the outcome in this game: the best separating equilibrium.

In signaling games with two types, this equilibrium dominance-based refinement is equivalent to the *intuitive criterion* proposed in Cho and Kreps (1987). Formally, a PBE is said to violate the intuitive criterion if there exists a type θ and an action $a \in A$ such that

$$\underset{s \in S^*(\Theta^{**}(a), a)}{\text{Min}} \; u_1(a, s, \theta) > u_1^*(\theta). \tag{13.AA.4}$$

Thus, we eliminate a PBE using the intuitive criterion if there is some type θ who has a deviation that is *assured* of yielding her a payoff above her equilibrium payoff as long as players $2, \ldots, I$ do not assign a positive probability to the deviation having been made by any type θ for whom this action is equilibrium dominated. We can think of the intuitive criterion as saying that to eliminate a PBE we must find a type of player 1 who wants to deviate even if she is not sure what exact belief of players $2, \ldots, I$ will result, she is only sure that they will not think she is a type who would find the deviation to be an equilibrium-dominated action. In general, the intuitive criterion is a more conservative elimination procedure than just insisting on PBEs involving reasonable beliefs using set $\Theta^{**}(a)$ because any PBE with reasonable beliefs using set $\Theta^{**}(a)$ passes the intuitive criterion's test, but as Example 13.AA.1 illustrates, a PBE could satisfy the intuitive criterion's test but fail to have reasonable beliefs. However, when there are only two types of player 1, the two notions are equivalent.

Example 13.AA.1: Suppose that there are three types of player 1, $\{\theta_1, \theta_2, \theta_3\}$, and that in some PBE the out-of-equilibrium action $\hat{a}$ is equilibrium dominated for type θ_1 only, so that $\Theta^{**}(\hat{a}) = \{\theta_2, \theta_3\}$. Suppose also that type θ_2 strictly prefers to deviate to action $\hat{a}$ if and only if beliefs over types θ_2 and θ_3 have $\mu(\theta_2 \mid \hat{a}) \geq \frac{1}{4}$ while type θ_3 strictly prefers to deviate to action $\hat{a}$ if and only if $\mu(\theta_2 \mid \hat{a}) \leq \frac{3}{4}$. This situation will not violate the intuitive criterion because condition (13.AA.4) does not hold for either type θ_2 or type θ_3. But in any PBE with reasonable beliefs using set $\Theta^{**}(a)$, one of the two types will deviate to action $\hat{a}$; therefore, this PBE must not have reasonable beliefs in this sense. When there are only two possible types for player 1, say θ_1 and θ_2, this difference disappears because whenever equilibrium domination eliminates a type from consideration, so that $\Theta^*(a) = \{\theta_i\}$ for $i = 1$ or 2, there is only one possible belief for players $2, \ldots, I$ to hold. ■

Although the use of either equilibrium domination or the intuitive criterion yields a unique prediction in the education signaling model when there are two types of workers, they do not accomplish this when there are three or more possible worker types (see Exercise 13.AA.1). Stronger refinements such as Banks and Sobel's (1987) notions of *divinity* and *universal divinity*, Cho and Kreps' (1987) related notion called *D1*, and Kohlberg and Mertens' (1986) *stability* do yield the unique prediction of the best separating equilibrium in these games with many worker types. See Cho and Kreps (1987) and Fudenberg and Tirole (1992) for further details.

REFERENCES

Akerlof, G. (1970). The market for lemons: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* **89**: 488–500.

Banks, J., and J. Sobel. (1987). Equilibrium selection in signaling games. *Econometrica* **55**: 647–62.

Cho, I-K., and D. M. Kreps. (1987). Signaling games and stable equilibria. *Quarterly Journal of Economics* **102**: 179–221.

Dasgupta, P., and E. Maskin. (1986). The existence of equilibrium in discontinuous economic games. *Review of Economic Studies* **46**: 1–41.

Fudenberg, D., and J. Tirole. (1992). *Game Theory*. Cambridge, Mass.: MIT Press.

Hellwig, M. (1986). Some recent developments in the theory of competition in markets with adverse selection. (University of Bonn, mimeographed).

Holmstrom, B., and R. B. Myerson. (1983). Efficient and durable decision rules with incomplete information. *Econometrica* **51**: 1799–819.

Kohlberg, E., and J.-F. Mertens. (1986). On the strategic stability of equilibria. *Econometrica* **54**: 1003–38.

Maskin, E., and J. Tirole. (1992). The principal-agent relationship with an informed principal, II: Common values. *Econometrica* **60**: 1–42.

Miyazaki, H. (1977). The rat race and internal labor markets. *Bell Journal of Economics* **8**: 394–418.

Riley, J. (1979). Informational equilibrium. *Econometrica* **47**: 331–59.

Rothschild, M., and J. E. Stiglitz. (1976). Equilibrium in competitive insurance markets: An essay in the economics of imperfect information. *Quarterly Journal of Economics* **80**: 629–49.

Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics* **87**: 355–74.

Spence, A. M. (1974). *Market Signaling*. Cambridge, Mass.: Harvard University Press.

Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory* **16**: 167–207.

Wilson, C. (1980). The nature of equilibrium in markets with adverse selection. *Bell Journal of Economics* **11**: 108–30.

EXERCISES

13.B.1[A] Consider three functions of $\hat{\theta}$: $r(\hat{\theta})$, $E[\theta \mid \theta \leq \hat{\theta}]$, and $\hat{\theta}$. Graph these three functions over the domain $[\underline{\theta}, \bar{\theta}]$, assuming that the first two functions are continuous in $\hat{\theta}$ but allowing them to be otherwise quite arbitrary. Identify the competitive equilibria of the adverse selection model of Section 13.B using this diagram. What about the Pareto optimal labor allocation? Now produce a diagram to depict each of the situations in Figures 13.B.1 to 13.B.3.

13.B.2[B] Suppose that $r(\cdot)$ is a continuous and strictly increasing function and that there exists $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ such that $r(\theta) > \theta$ for $\theta > \hat{\theta}$ and $r(\theta) < \theta$ for $\theta < \hat{\theta}$. Let the density of workers of type θ be $f(\theta)$, with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. Show that a competitive equilibrium with unobservable worker types necessarily involves a Pareto inefficient outcome.

13.B.3[B] Consider a *positive selection* version of the model discussed in Section 13.B in which $r(\cdot)$ is a continuous, strictly *decreasing* function of θ. Let the density of workers of type θ be $f(\theta)$, with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$.

(a) Show that the *more capable* workers are the ones choosing to work at any given wage.

(b) Show that if $r(\theta) > \theta$ for all θ, then the resulting competitive equilibrium is Pareto efficient.

(c) Suppose that there exists a $\hat{\theta}$ such that $r(\theta) < \theta$ for $\theta > \hat{\theta}$ and $r(\theta) > \theta$ for $\theta < \hat{\theta}$. Show that any competitive equilibrium with strictly positive employment necessarily involves *too much* employment relative to the Pareto optimal allocation of workers.

13.B.4[B] Suppose two individuals, 1 and 2, are considering a trade at price p of an asset that they both use only as a store of wealth. Ms. 1 is currently the owner. Each individual i has a privately observed signal of the asset's worth y_i. In addition, each cares only about the expected value of the asset one year from now. Assume that a trade at price p takes place only if both parties think they are being made strictly better off. Prove that the probability of trade occurring is zero. [*Hint*: Study the following trading game: The two individuals simultaneously say either "trade" or "no trade," and a trade at price p takes place only if they both say "trade."]

13.B.5[B] Reconsider the case where $r(\theta) = r$ for all θ, but now assume that when the wage is such that no workers are accepting employment firms believe that any worker who might accept would be of the lowest quality, that is, $E[\theta \mid \Theta = \varnothing] = \underline{\theta}$. Maintain the assumption that all workers accept employment when indifferent.

(a) Argue that when $E[\theta] \geq r > \underline{\theta}$, there are now two competitive equilibria: one with $w^* = E[\theta]$ and $\Theta^* = [\underline{\theta}, \bar{\theta}]$ and one with $w^* = \underline{\theta}$ and $\Theta^* = \varnothing$. Also show that when $\underline{\theta} \geq r$ the unique competitive equilibrium is $w^* = E[\theta]$ and $\Theta^* = [\underline{\theta}, \bar{\theta}]$, and when $r > E[\theta]$ the unique competitive equilibrium is $w^* = \underline{\theta}$ and $\Theta^* = \varnothing$.

(b) Show that when $E[\theta] > r$ and there are two equilibria, the full-employment equilibrium Pareto dominates the no-employment one.

(c) Argue that when $E[\theta] \geq r$ the unique SPNE of the game-theoretic model in which two firms simultaneously make wage offers is the competitive equilibrium when this equilibrium is unique, and is the full-employment (highest-wage) competitive equilibrium when the competitive equilibrium is not unique and $E[\theta] > r$. What happens when $E[\theta] = r$? What about the case where $E[\theta] < r$?

(d) Argue that the highest-wage competitive equilibrium is a constrained Pareto optimum.

13.B.6[C] [Based on Wilson (1980)] Consider the following change in the adverse selection model of Section 13.B. Now there are N firms, each of which wants to hire at most 1 worker. The N firms differ in their productivity: In a firm of type γ a worker of type θ produces $\gamma\theta$ units of output. The parameter γ is distributed with density function $g(\cdot)$ on $[0, \infty]$, and $g(\gamma) > 0$ for all $\gamma \in [0, \infty]$.

(a) Let $z(w, \mu)$ denote the aggregate demand for labor when the wage is w and the average productivity of workers accepting employment at that wage is μ. Derive an expression for this function in terms of the density function $g(\cdot)$.

(b) Let $\mu(w) = E[\theta \mid r(\theta) \leq w]$, and define the *aggregate demand function for labor* by $z^*(w) = z(w, \mu(w))$. Show that $z^*(w)$ is strictly increasing in w at wage $\bar{w}$ if and only if the elasticity of μ with respect to w exceeds 1 at wage $\bar{w}$ (assume that all relevant functions are differentiable).

(c) Let $s(w) = \int_{\underline{\theta}}^{r^{-1}(w)} f(\theta)\, d\theta$ denote the *aggregate supply function of labor*, and define a competitive equilibrium wage w^* as one where $z^*(w^*) = s(w^*)$. Show that if there are multiple competitive equilibria, then the one with the highest wage Pareto dominates all the others.

(d) Consider a game-theoretic model in which the firms make simultaneous wage offers, and denote the highest competitive equilibrium wage by w^*. Show that (i) only the highest-wage competitive equilibrium can arise as an SPNE, and (ii) the highest-wage competitive equilibrium is an SPNE if and only if $z^*(w) \leq z^*(w^*)$ for all $w > w^*$.

13.B.7[B] Suppose that it is impossible to observe worker types and consider a competitive equilibrium with wage rate w^*. Show that there is a Pareto-improving market intervention $(\tilde{w}_e, \tilde{w}_u)$ that reduces employment if and only if there is one of the form $(w_e, w_u) = (w^*, \hat{w}_u)$ with $\hat{w}_u > 0$. Similarly, argue that there is a Pareto-improving market intervention $(\tilde{w}_e, \tilde{w}_u)$ that increases employment if and only if there is one of the form $(w_e, w_u) = (\hat{w}_e, 0)$ with $\hat{w}_e > w^*$. Can you use these facts to give a simple proof of Proposition 13.B.2?

13.B.8[B] Consider the following alteration to the adverse selection model in Section 13.B. Imagine that when workers engage in home production, they use product x. Suppose that the amount consumed is related to a worker's type, with the relation given by the increasing function $x(\theta)$. Show that if a central authority can observe purchases of good x but not worker types, then there is a market intervention that results in a Pareto improvement even if the market is at the highest-wage competitive equilibrium.

13.B.9[B] Consider a model of *positive selection* in which $r(\cdot)$ is strictly decreasing and there are two types of workers, θ_H and θ_L, with $\infty > \theta_H > \theta_L > 0$. Let $\lambda = \text{Prob}\,(\theta = \theta_H) \in (0, 1)$. Assume that $r(\theta_H) < \theta_H$ and that $r(\theta_L) > \theta_L$. Show that the highest-wage competitive equilibrium need not be a constrained Pareto optimum. [*Hint*: Consider introducing a small unemployment benefit for a case in which $E[\theta] = r(\theta_L)$. Can you use the result in Exercise 13.B.7 to give an exact condition for when a competitive equilibrium involving full employment is a constrained Pareto optimum?]

13.B.10[B] Show that Proposition 13.B.2 continues to hold when $r(\theta) > \theta$ for some θ.

13.C.1[B] Consider a game in which, first, nature draws a worker's type from some continuous distribution on $[\underline{\theta}, \bar{\theta}]$. Once the worker observes her type, she can choose whether to submit to a costless test that reveals her ability perfectly. Finally, after observing whether the worker has taken the test and its outcome if she has, two firms bid for the worker's services. Prove that in any subgame perfect Nash equilibrium of this model all worker types submit to the test, and firms offer a wage no greater than $\underline{\theta}$ to any worker not doing so.

13.C.2[C] Reconsider the two-type signaling model with $r(\theta_L) = r(\theta_H) = 0$, assuming a worker's productivity is $\theta(1 + \mu e)$ with $\mu > 0$. Identify the separating and pooling perfect Bayesian equilibria, and relate them to the perfect information competitive outcome.

13.C.3[B] In text.

13.C.4[B] Reconsider the signaling model discussed in Section 13.C, now assuming that worker types are drawn from the interval $[\underline{\theta}, \bar{\theta}]$ with a density function $f(\theta)$ that is strictly positive everywhere on this interval. Let the cost function be $c(e, \theta) = (e^2/\theta)$. Derive the (unique) perfect Bayesian equilibrium.

13.C.5[B] Assume a single firm and a single consumer. The firm's product may be either high or low quality and is of high quality with probability λ. The consumer cannot observe quality before purchase and is risk neutral. The consumer's valuation of a high-quality product is v_H; her valuation of a low-quality product is v_L. The costs of production for high (H) and low (L) quality are c_H and c_L, respectively. The consumer desires at most one unit of the product. Finally, the firm's price is regulated and is set at p. Assume that $v_H > p > v_L > c_H > c_L$.

(a) Given the level of p, under what conditions will the consumer buy the product?

(b) Suppose that before the consumer decides whether to buy, the firm (which knows its type) can advertise. Advertising conveys no information directly, but consumers can observe the total amount of money that the firm is spending on advertising, denoted by A. Can there be a separating perfect Bayesian equilibrium, that is, an equilibrium in which the consumer rationally expects firms with different quality levels to pick different levels of advertising?

13.C.6[C] Consider a market for loans to finance investment projects. All investment projects require an outlay of 1 dollar. There are two types of projects: good and bad. A good project has a probability of p_G of yielding profits of $\Pi > 0$ and a probability $(1 - p_G)$ of yielding profits of zero. For a bad project, the relative probabilities are p_B and $(1 - p_B)$, respectively, where $p_G > p_B$. The fraction of projects that are good is $\lambda \in (0, 1)$.

Entrepreneurs go to banks to borrow the cash to make the initial outlay (assume for now that they borrow the entire amount). A loan contract specifies an amount R that is supposed to be repaid to the bank. Entrepreneurs know the type of project they have, but the banks do not. In the event that a project yields profits of zero, the entrepreneur defaults on her loan contract, and the bank receives nothing. Banks are competitive and risk neutral. The risk-free rate of interest (the rate the banks pay to borrow funds) is r. Assume that

$$p_G\Pi - (1 + r) > 0 > p_B\Pi - (1 + r).$$

(a) Find the equilibrium level of R and the set of projects financed. How does this depend on p_G, p_B, λ, Π, and r?

(b) Now suppose that the entrepreneur can offer to contribute some fraction x of the 1 dollar initial outlay from her own funds ($x \in [0, 1]$). The entrepreneur is liquidity constrained, however, so that the effective cost of doing so is $(1 + \rho)x$, where $\rho > r$.

(i) What is an entrepreneur's payoff as a function of her project type, her loan-repayment amount R, and her contribution x?

(ii) Describe the best (from a welfare perspective) separating perfect Bayesian equilibrium of a game in which the entrepreneur first makes an offer that specifies the level of x she is willing to put into a project, banks then respond by making offers specifying the level of R they would require, and finally the entrepreneur accepts a bank's offer or decides not to go ahead with the project. How does the amount contributed by entrepreneurs with good projects change with small changes in p_B, p_G, λ, Π, and r?

(iii) How do the two types of entrepreneurs do in the separating equilibrium of **(b)**(ii) compared with the equilibrium in **(a)**?

13.D.1[B] Extend the screening model to a case in which tasks are productive. Assume that a type θ worker produces $\theta(1 + \mu t)$ units of output when her task level is t where $\mu > 0$. Identify the subgame perfect Nash equilibria of this model.

13.D.2[B] Consider the following model of the insurance market. There are two types of individuals: high risk and low risk. Each starts with initial wealth W but has a chance that an accident (e.g., a fire) will reduce her wealth by L. The probability of this happening is p_L for low-risk types and p_H for high-risk types, where $p_H > p_L$. Both types are expected utility maximizers with a Bernoulli utility function over wealth of $u(w)$, with $u'(w) > 0$ and $u''(w) < 0$ at all w. There are two risk-neutral insurance companies. An insurance policy consists of a premium payment M made by the insured individual to her insurance firm and a payment R from the insurance company to the insured individual in the event of a loss.

(a) Suppose that individuals are prohibited from buying more than one insurance policy. Argue that a policy can be thought of as specifying the wealth levels of the insured individual in the two states "no loss" and "loss."

(b) Assume that the insurance companies simultaneously offer policies; as in Section 13.D, they can each offer any finite number of policies. What are the subgame perfect Nash equilibrium outcomes of the model? Does an equilibrium necessarily exist?

13.D.3[C] Consider the following extension of the model you developed in Exercise 13.D.1. Suppose that there is a fixed task level T that all workers face. The monetary equivalent cost of accepting employment at this task level is $c > 0$, which is independent of worker type. However, now a worker's actual output is observable and verifiable, and so contracts can base compensation on the worker's ex post observed output level.

(a) What is the subgame perfect Nash equilibrium outcome of this model?

(b) Now suppose that the output realization is random. It can be either good (q_G) or bad (q_B). The probability that it is good is p_H for a high-ability worker and p_L for a low-ability worker ($p_H > p_L$). If workers are risk-neutral expected utility maximizers with a Bernoulli utility function over wealth of $u(w) = w$, what is the subgame perfect Nash equilibrium outcome?

(c) What if workers are strictly risk averse with $u''(w) < 0$ at all w?

13.D.4[B] Reconsider the screening model in Section 13.D, but assume that (i) there is an infinite number of firms that could potentially enter the industry and (ii) firms can each offer at most one contract. [The implication of (i) is that, in any SPNE, no firm can have a profitable entry opportunity.] Characterize the equilibria for this case.

13.AA.1[C] Consider the extension of the signaling model discussed in Section 13.C to the case of three types. Assume all three types have $r(\theta) = 0$. Provide an example in which more than one perfect Bayesian equilibrium satisfies the intuitive criterion.